



TITLE  
JOURNAL

Direct Submission  
Submitted (08-MAY-1998) Pathology, University of Texas Medical  
Branch at Galveston, Keller 2.138, 301 University Blvd, Galveston,  
TX 77555 -0609, USA

## FEATURES

## SOURCE

Location/Qualifiers

1..2500

/organism="Plasmodium gallinaceum"

/db\_xref="taxon:5849"

/dev\_stage="ookinete"

276..2039

/EC\_number="3.2.1.14"

/codon\_start=1

/product="endochitinase precursor"

/protein\_id="AA063208.1"

/db\_xref="GI:7530424"

/translation="MNFKISFLIVLSVANSRPTLKGNKNNISLGIIEKNKTH  
OTIEHSEFSHKSNNSNFEVSGYCGGCSNRITKNNKINNDKRSFRLLEKKR  
KQIITACIYSSNSQGRKHMIDSNPWSLILYAFRIIMLVDSFNRQFLLR  
KHLEIETYSMLNIRIRKRPDVIILSLIGETIMIDEKEIDYDKTLKLVNP  
DLGVDIDMPEHGFKNLNLNSNYIKILNLRKTIPEEKLSISGNSAALSCVS  
GVASFCDSESPYNTKRLSIOETNKLHRAAAMLNAGTFINFTNKKEDIVFIOT  
YLNETNPDIIMVYLISHLYEGLKYNTIILGSLNNRGGFSPEKLELVEKTIH  
DKNNNNRAGISCIIMHLPKKEQLPTGSDVDITLNTMKHLNPEVQPKDITENE  
DCSTIDIEYFGLVPIITGIYKHNDAIMKRSYSIAHGVDRMVDLVKCYEKICDG  
KAHHYNTDKRESSIIIMKEPYLIKMMOOGPEGQALLESYTKLDASKCPGIEEMNK  
YPRKPLEVEYOEVDLPLO"

BASE COUNT 993 a 283 c 365 g 858 t 1 others  
ORIGIN

Query Match 100.0%; Score 1764; DB 3; Length 2500;  
Best Local Similarity 100.0%; Pred. No. 2.4e-242;  
Matches 1764; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaatttaataatcaatattttaaattatagatcatcattcgtatttcgaattcc 60  
DB 276 ATGAATTTTAAATATCAATATTTTAAATATGATATCCATCTGTATTCGCAAAATTC 335  
QY 61 agaaccttgaagaagaaataataataatcaatcattcgggaataatcaggaataaa 120  
DB 336 AGAACCTTGAAGAAATAATATATTAATTCATTTGGGAATTAACGGGAAATTA 395  
QY 121 aataaacatcatcaaacggaataacatgagctctttcacacttaacatcgaataagt 180  
DB 396 AATAAATCATCAACGGAATAATCATGAGCTTTTTCACATCTTAATGCAATTAATAGT 455  
QY 181 aatttgtagaataatgatacttatttcggagatgggtgttaacctggaattacaaaaat 240  
DB 456 AATTTTGTGAAATATGATCTTATTCGCGAGATGGGTGTAACCTGAAATTAACAAAAAT 515  
QY 241 aataaaaataataataaaatgataagaatacccaagacaaattttaaggaggatataa 300  
DB 516 AATAAATAATATTAATTAATGATAGCAAAATCCCAAGCAAAATTTAGAGGAGTATAA 575  
QY 301 aaaaggaacaagaatattatagcaggaatacctatggtltaatggaacagtcaggtaga 360  
DB 576 AAAAGGAAACAAGGATTAATAGCAGATCTATGCTTCAATGGAACAGTCAGAGTGATGA 635  
QY 361 gcaaacatatgattcacaaccaatggtgtcaattttatataatgcatctgcgc 420  
DB 636 GCAAAACATATGATGATTCACCAACCAATGGTCAATTTTAATATTCATTTGCTCCG 695  
QY 421 attaatatgatatgatactatagacacatttaatggaagcaaaatgctcattaaaga 480  
DB 696 ATTATATATGATATGATGATCTAGACCAATTAATGGAAGCAAAAGATTCCTATTAA 755  
QY 481 aaacaaggcttagaataagaaccatgtagatgcttaataagaattagacgtatcaga 540  
DB 756 AAACAAGGCTTAGAATAAGAAACCATGATGATGCTTAATGAATTAAGACGTATCAGA 815  
QY 541 aaagtagcagatgtaattattctttatcctaggttggagaaacccatataatgataat 600  
DB 816 AAAGTAGCTCAGATGATTAATTTCTTTATCCCTTAAGGTGGAAGCAACCATTAATGATAGAT 875

QY 601 atagaaaaagaattgattatgttgataaataattgaagcttgtaatgattttgattta 660  
DB 876 ATAGAAAAAGAATTGATTATGATGATTAATAATATGAAGCTGTTAATGATTGATTTA 935  
QY 661 gatgtgtgataatgactcgggaaccacatggaaggtttcaacttaaatgaaataat 720  
DB 936 GATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995  
QY 721 ttccaattatataataataatgataatgataatgataatgataatgataatgataatg 780  
DB 996 TTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1055  
QY 781 ttaattcaattctgtgtatcaaatgctgataatcaatgctgttcaaggagttgacat 840  
DB 1056 TTAATTTCAATTTCTGTTTCAATTAATGATGATGATGATGATGATGATGATGATGAT 1115  
QY 841 ttctgtaagtgtaagaatccatcaataacataatcttctgtgataaataagaaca 900  
DB 1116 TTCTGTAAGATGAAGAATTCCTCATTAATTAATTAATTAATTAATTAATTAATTAAT 1175  
QY 901 aataaagaattacatagagcagcagatgtatcagcaggaactttatataatttt 960  
DB 1176 AATAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1235  
QY 961 aatacagaagaaggaataatagatctgtatatttcaataacatacaatltagaactaca 1020  
DB 1236 AATACAGCAAGGAAAGAAATAGATCTGTATTAATTAATTAATTAATTAATTAATTAAT 1295  
QY 1021 aatcagataatgataatgataatgataatgataatgataatgataatgataatgata 1080  
DB 1296 AATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1355  
QY 1081 atcaacatcatataggtttttcatagaaacatacagagtgattgattccggaat 1140  
DB 1356 ATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1415  
QY 1141 aaagaattatagaaattggttagaagaacaatacatgataaatacaaaaataatag 1200  
DB 1416 AAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1475  
QY 1201 gcaagtgataaggaataggaatgataatgataatgataatgataatgataatgata 1260  
DB 1476 GCAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1535  
QY 1261 gatgtagataatttcttcaacaatatlttgaaacatttaacccgaagttaacaactca 1320  
DB 1536 GATGTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1595  
QY 1321 aaagaccttaactaactgaaacccctgaagactgtagacaataatgataatgttcca 1380  
DB 1596 AAAGACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1655  
QY 1381 ggaactgtatcccaacataaggaatataatgataatgataatgataatgataatgata 1440  
DB 1656 GGACTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1715  
QY 1441 agacttattcaatcaatgacactgtgttagacagataatgaatggaactgtgcaagta 1500  
DB 1716 AGACTTATTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1775  
QY 1501 tgcattgaaaaataatgagatggaagaagccacttatttaacactgactataaaga 1560  
DB 1776 TGCATTGAAAAATAATGAGATGGAAGACAGCCATTTATTAACACGATTAATTAAGAA 1835  
QY 1561 agctctattatataatggaagaaggaacataatgataatgataatgataatgataatg 1620  
DB 1836 AGCTCTATTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1895  
QY 1621 ccggaaggtcaggaactagatcatatacaaaatagatgcatcaaatgtccaggata 1680  
DB 1896 CCGGAAGGTCCAGGACCTAGAGATCATACCAAACTAGATGATCATCAATGTCCAGGATA 1955

FEATURES	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1681	gaagatggaataaataatccacaataaccactcagaagatgaggaacataatgacaa	1740													
1936	GAAAGATGGAATGAAATGAAATATTCACATTAACCACTAGAGAGAGAAACAATATGACAA	2015													
1741	gaagtcgattaccataataa	1764													
2016	GAAATGGAATTTTACCTTTACATTA	2039													
AF216703	1764 bp	DNA	SYN	26-JUN-2000											
AF216703	Synthetic construct chitinase CHT1 (Chl1) gene, complete cds.														
AF216703.1	GI:8705223														
	synthetic construct.														
	artificial sequence.														
1 (bases 1 to 1764)															
Vinetz,J.M.															
Chitinases of the avian malaria parasite, Plasmodium gallinaceum															
Unpublished															
2 (bases 1 to 1764)															
Vinetz,J.M.															
Direct Submission															
Submitted (14-DEC-1999) WHO Center for Tropical Disease, University of Texas Medical Branch, 301 University Blvd., Galveston, TX															
77555-0609, USA															
Location/Qualifiers															
1. 1764															
/organism="synthetic construct"															
/db_xref="taxon:32630"															
/note="Plasmodium gallinaceum gene engineered for bacterial codon usage"															
1. 1764															
/gene="Chl1"															
1. 1764															
/gene="Chl1"															
/codon_start=1															
/transl_table=1															
/product="chitinase CHT1"															
/protein_id="AAF8777.1"															
/db_xref="GI:8705224"															
/translation="MNFKISFLIIVSILYSANSRSLGKNNINNSLGIIRENKNRTH															
QTEHSEFSLKSNNSNFEVSGYCGCSRTKNNKNNKNDKSPQILEEYKKR															
KCGIAGYVGSNMGDRAKHMIDSNMVSILYIAFARIMLVDVSRPGRPLLR															
KHGIEFTYGMMLNEIRIKRYPDVIILSLGERTMIDIEKIDYVDKILAVDF															
DLDVDIDMEHGFNLFNLSNYIKILINLRKTIPEKLIISGSSNALSCVS															
GVASFCDDESPYNTKFLSQIETNKLHAAAMLSSGPIINFENAKELIDVFQI															
YNETNPDIADMDVYLSHLYFGKLYNTIILGFSLEHNRG															

D	b	121	AACAAACCCTCAAAACCGAATAATC	ATCATAGTCCCTTTCTTCTCACTTAAATCAATATATAGC	180
O	y	181	aattcttgtagaataatgatacttaatt	ctgaggagatggggtglaactctagaattacaaaaat	240
D	b	181	TAATTCGTTGAGTAGCGTAGCTAT	GTGGCATGGCTTAACACCCGATACACCAAAAC	240
O	y	241	aataaataataataaaaaatgata	tgagaaatccacaagacaaaatttttagagatataa	300
D	b	241	AACAAAACATCAATAAATAAACG	ACGCAAAAGCCCGCCAGATTCCTGAAGAATATCAAA	300
O	y	301	aaaaggaacaagatattatagca	gaactatggtttcatagtaacaagtcacaagtgataga	360
D	b	301	AAAGCCAAACAGGCGATTAT	TCCGGGCTACTATGGTTCTCTGGAAACGCGACGCGACCGT	360
O	y	361	gcaaaacataatgattgatacca	aaoccaaagtggtgtlcaatttatalatgcatltgtctgc	420
D	b	361	GCCAAACACATGATGATGAC	CTTAATCCGATGGTGCATCCTGTATATGCTTGCCCGCT	420
O	y	421	attaatagtataatgatagtat	ctcagaccatttaagtgaagacaaagattcctattaaga	480
D	b	421	ATTATATAGCGTAGAGCGTT	CTCGCTGTTTAAAGGTGCCACGCGTTTCTCTCGTGGT	480
O	y	481	aaacccgcttagaataatgaa	aacctaagttatgatacttaagaaatagacgtatacga	540
D	b	481	AAGCATGGCCCTTGAGTAC	GAACACCTATGTGCATGATGCTGAATGTGAATTCGTGCGCATCCG	540
O	y	541	aaagtcagtlccagatgtaata	ttattcttataccttaagtgagaaacctaatalatagat	600
D	b	541	AAAGTACCCCGGAGAGTAA	TGATCTGTTTAAAGCTTAGCGGCGCAAAACCTATATGATGAC	600
O	y	601	atagaaaaaagaattgata	tggtgataataatattgaagcttgtttaagatttggattta	660
D	b	601	ATCGAAAAAGAAATGCA	TATGTCCATTAATAATTCGAAACATGAGTAAGATTTTGATCTG	660
O	y	661	gatggtgtagatatgtact	ctgggaaccacataggaagtttacaactaaagaaat	720
D	b	661	GATGGTGTGACATGAC	TGACTGGGAACCGCATGGCAAGTTCTACATCTGTAATGAACTGAC	720
O	y	721	tttccaattatlatatlaa	atlaaacttgbttaaagaaaaactatccggaagaanaag	780
D	b	721	TTTCACTAATTAATATCA	AACTGATGTATTAACCTGCTGGCGCAAAACCATCCGAGGAGAAAA	780
O	y	781	ttaattcaattctcgtgt	ctatcaaatatgcgtcatataatgcgtttcaggagttgcactc	840
D	b	781	CTGATCTTATCTCTCGTAG	CAGCAACACACACATGACCTCGTTAGCGGCGTGTCT	840
O	y	841	ttctgtlaaagatgaaga	atccatcaaaactaaatttggcttgaaacaaatagaaca	900
D	b	841	TTTGTCAAAAGATGAAGA	ATCCCGGTATTAACCTTAATTCCTGAGGACAAATGAAACT	900
O	y	901	aataaagaatfatcatag	ggcagcagctgtlatatcagaggaacttttataattttt	960
D	b	901	AACAAAGAGCTGACAT	CGTGTGCGCCATGCTGAGGCGGGTACGTTTCAACTTTTTC	960
O	y	961	aatacagcaagaggaaaa	tagatctctgtlatatttcaacaatactatagaactga	1020
D	b	961	AACACTGCGAAAGAAAAA	ATGACCTGTGTTATTTCAAAACCTTAATCTGAGACACACC	1020
O	y	1021	aatcagaataatagtgata	tgataatgactatccacttaatttttggtttaaataatac	1080
D	b	1021	AACCGGATATTATGCTG	ATATGTATGTGAGCCACCTGTATTTTCGGCTTAATAATACAC	1080
O	y	1081	atcacaaatcatataggt	tttctcatatagaacatacagaagtgatagcccgaaat	1140
D	b	1081	ATTACGATCATTCCTG	GGTTTTCTCTGTGAAACAAATCCTGGCGGTTTCAGCCCGGAGAAC	1140
O	y	1141	aaagaattatagaattg	gttagaaaaaacaatagataaanaalcaaaataataatagg	1200
D	b	1141	AAAGAACTGCTGGAAT	TGTTGGCAAAACCTATCCACGACAGAACCAAGATAATATACCGT	1200
O	y	1201	gcagatggtatagggata	ctggcatatttatagaagaacaataaccactggtatctc	1260
D	b	1201	GCTGATGTATTCGAT	TATTTTGGCATCTGTTTCAATGAAGAAACAATGCCACCTGTAGACTTT	1260

QY 1261 gatgtagatattttcttacaatatttgaaacaattaaactcctgaagtcacaactca 1320  
DB 1261 GAGCTGATATTTTTCGACGACATTTGAAACATCGAACCCGAGTGCACAAACCCCG 1320  
QY 1321 aaagacctataactgaagaacccggaagcgttagcacaatagatgaatagtcca 1380  
DB 1321 AAGACTTAACCATACCGGAAACCCGGAAGACTGCTCCAGATCGAGAGTATGTACCG 1380  
QY 1381 ggaactgtatltccaacataagagatataatacaacaatga tgcataatgaagaact 1440  
DB 1381 GGTCTGTGTATTCGACATCGCATCTACTATAACACACAGATCGCATTTGGAAACC 1440  
QY 1441 agacttatcaatcaatgaaccgtgtgtacagacatgaatgagactgttcaaga 1500  
DB 1441 CGTTCTACAGCATCCATGACCGGCGTGCATCGACGATGCGGATCTGGTGAAGCTT 1500  
QY 1501 tgcataaataatcgaatggaagacagccatattataaactgaactgaagaaga 1560  
DB 1501 TGCTATGAGAAATTTGTGATGGCAAGCTGCGCATTTATTAATACCGACTATTAAGAA 1560  
QY 1561 agctctattatataatgaagaaggaacacatatttaataatggtgcaacaagaact 1620  
DB 1561 AGCTCTATTATCATTTGAGAGGCTGACCGTACTTATCAAAATGCTGACGACAGGCCCG 1620  
QY 1621 ccggaaggtcagcactagatagatcacacaactagaatgcacccaagtgcaggata 1680  
DB 1621 CCGAAGGTTCAGCCCTGAGACTACACCAACTGATCTCTCAAGTCTCCGGCATTT 1680  
QY 1681 gaagaatgaataaataatccacaataacactagaagtagagagaatgaagaaga 1740  
DB 1681 GAAGATGGAACAGAAAGATATCGGCAACACCCCTGGAAGTGAAGACAGTACGACACG 1740  
QY 1741 gaagtgatattaccatcaataa 1764  
DB 1741 GAAGTGACCTGCGCTGCATTA 1764

RESULT 3  
PBE305256  
LOCUS PBE305256 2562 bp mRNA INV 30-MAY-2001  
DEFINITION Plasmodium berghei mRNA for chitinase (cht1 gene).  
ACCESSION AJ305256  
VERSION AJ305256.1 GI:14275848  
KEYWORDS chitinase; CHT1 gene.  
SOURCE Plasmodium berghei.  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 2562)  
AUTHORS Dessens J.T., Mendoza J., Claudianos C., Vlnetz J.M., Khater E., Hassard S., Ranawake G.R. and Sinden R.E.  
TITLE Knockout of the rodent malaria parasite chitinase pbcht1 reduces infectivity to mosquitoes  
JOURNAL Infect. Immun. 69 (5), 4041-4047 (2001)  
MEDLINE 21246711  
REFERENCE 2 (bases 1 to 2562)  
AUTHORS Dessens J.T.  
TITLE Direct Submission  
JOURNAL Submitted (31-JAN-2001) Dessens J.T., Dept. of Biology, Imperial College of Science, Technology and Medicine, SAFB, Imperial College Road, London, SW7 2AZ, UNITED KINGDOM  
FEATURES  
source  
1. 2562  
/organism="Plasmodium berghei"  
/strain="ANKA"  
/isolate="2.34"  
/db\_xref="taxon:5821"  
174..227  
/gene="cht1"  
174..2120  
/gene="cht1"  
/function="mosquito infection"  
/codon\_start=1

/product="chitinase"  
/protein\_id="CAC40151.1"  
/db\_xref="GI:14275849"  
/translation="MNIVYSLPLFISTJNNAVSVGLKGGTTINPINSHEKOKYS  
LSHVDAICGSKRYVCHETKICNGSKGERCEGSGNGAENGNGVENGNGVENG  
GAENGNGVENGNGVENGNGAENGNGVENGNGVENGNGVENGNGVENGNGVENG  
EYKRRKOGIAGYYSVNSOGASGOMTHSNHVSIIYIARIMNVDASKPYNGL  
OKFLORHGLEFETYGMAANEIKRIKARPVDVILISLGETYIMIDITREIDYIPOIV  
NMVNFNLDGVDIDMEPNKGFYHLNLFANSYIKLIDLSINISKDKLISGSSNA  
ALSCVSWERPECKDEETPNYTOYLSPOIMVNBELTKASYSLSGTFRVYENARKID  
LVFIORYNETSNPDIMDMYLSHLFPGKRYDITFLGISTONKGSPPSKNIVEL  
VSKATIDNNHEFRADGVGIMHLEKQMPNCAVYDIDGFDIWTWHLNPGKLSKPDIV  
SIPNPDSCSTIEEYVGVIPVGYTAKINDAIWYTRSYTRAPIDREEMLVVVCY  
EKAKNNEAHYVNDVPTGCTKIIMGEVFSIKOMHGPEGPYLESEKIDEXSOPGL  
AEWNIRHPKPIEVEAPYQELDQ"  
174..2120  
gene  
mat\_peptide  
228..2117  
/gene="cht1"  
/product="chitinase"  
BASE COUNT 1001 a 323 c 416 g 822 t  
ORIGIN  
Query Match 53.1%; Score 937.2; DB 3; Length 2562;  
Best Local Similarity 72.7%; Pred. No. 1.1e-124;  
Matches 1225; Conservative 0; Mismatches 458; Indels 3; Gaps 1;  
QY 79 aataataataataatcatattggaataataatcgggaataataataataatcaataacg 138  
DB 441 AATGGCTGTAAATAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGCT 500  
QY 139 gaatacatgagctctttccacatcttaaatcgaataatgaatattgtagaataatga 198  
DB 501 GCAGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 560  
QY 199 tct---tatgagagatgggtgtaactagaatcaaaaaataataataataat 255  
DB 561 AATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 620  
QY 256 aaaaatgataaataatcacaagaacaatttagagagatataaaaaagaagaagaagt 315  
DB 621 AACCATGAAAGAAATCTCCAGGGAATTTTGAAGATATTAAGAAAGAAAGCAAGCT 680  
QY 316 attatagcagatcattatgattcattatgaaacagtcgaatgagacaaacataatgatt 375  
DB 681 ATAAATGCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 740  
QY 376 gatcaaacccaatggtgtcaatttataatattgaattgtgcgaataatattgattat 435  
DB 741 CATTCACACCCACACGATCTATTAATTTATATAGCTTTCGCCCGGTATTAATGAACTAT 800  
QY 436 gatgatcagaccattatgaagaagaagaatctcattatgaagaagaagaagaagtga 495  
DB 801 GATGCTTTAAACCATATATATGACTTCAAAATTTTTCACAAAACCAATGACTAGAA 860  
QY 496 tatgaaacctatgattatgattatgaatgaatgaatgaatgaatgaatgaatgaatga 555  
DB 861 TTTGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 920  
QY 556 gtaattatcttattccttaagtgaggaagaacataatgataatgaagaagaat 615  
DB 921 GTAATTAATTAATTTTATCATTTAGAGAGAAACATCATGATATGATATGAAAGAAAT 980  
QY 616 gattatgagataaataatgaagctgtgataatgattttagatgagtgatgataat 675  
DB 981 GATTATATTTCCCAATATGTAACATGCTTAACATTTTGAATGATGATGATGATGAT 1040  
QY 676 gacttggaacacatgaggaagttttacacactaaatgaatgaatgaatgaatgaatga 735  
DB 1041 GATTGGAAACCAATGGAAGATTTTACATTTAATGATGATGATGATGATGATGATGAT 1100  
QY 736 ataaataataatcattgtttaagaanaactatccggaagaagaagaatgaatgaatcatt 795



```

Db 1101 ATTAATGATTGATTATTAATAGAACACATTTCTAAGATAGCTAATTTCCATATCT 1160
QY 796 gttcacaacatgctgcatlcatcgcgtttcaggagttgcatcttctcgtcaagaatgaa 855
Db 1161 GGATCATCAATAGCAGATTTATCATGTGTATCATGATGAGGACATTTTGTAAAGATGAA 1220
QY 856 gaatccatataaacactaaatttctcgtcaacaatagaacaataaagaattacat 915
Db 1221 GAAAGCCATATTAATACAGATTTATCTGAAACAAATGATCGTAAACCCGATTTATAC 1280
QY 916 agggcagcagagatgtatcagcagacatttataatttataacagcagaagag 975
Db 1281 AAACATCAACCATGTTATCTACAGAACATTTGTTAATGTTTAAATACAGCAAGAT 1340
QY 976 aaaaatacctgtatattatcaacaacatacaattagaacaataccagataatg 1035
Db 1341 AAAATGATCTGTTTATTCAAACATATTAATTTAGAACATCAACCCGATATCATG 1400
QY 1036 gtatagatgtaactatcccatattatgttggttaaaataaacatcaacatata 1095
Db 1401 CTGTATATGATTTATCTCATTTTGTGAAAAAAATGTGATTAATGTTTCTTA 1460
QY 1096 gatttcattagacacatacagaggtgatttagtcccgcaaaataagaatitaga 1155
Db 1461 GGTTTTCATAGAACAAATAGAGTGATTTAGTCCATACACAAAACTTATGTA 1520
QY 1156 ttggtgaaaaaacaatatacgtataaaataaataaataagggcagatgtagatg 1215
Db 1521 TTGATATGCAAGCTATACATGACATATACGAAATTCACACAGCGAGTGTA 1580
QY 1216 atatgcatattatatagaagaacaattaccacgtgatactcgatgtagataat 1275
Db 1581 ATTTGGATTTATTAAGAAAGAAATGCTAATGAGCATATGATCTGATGTTT 1640
QY 1276 ctcaaatcttggaacatttaacatcctgaagtagacaacataaagaagactacta 1335
Db 1641 ATTGATATACATGACCCATTTAAATCCAGAAATTAATCCACACGATATGTTCA 1700
QY 1336 actgaaacctggaagactgtagacacatagatgaatagtctcgaactcgatlatca 1395
Db 1701 ATCCCTAACCTGACGATTTGATGACTATAGAACATATGATACAGAGTGTATCCA 1760
QY 1396 accatagggatatactacaacacacatgatactatagaactagatcttacaat 1455
Db 1761 ACTGTAGGACATATGCTAAACACAAATGATGCTATATGGGTAAACAGGCTCTATTCA 1820
QY 1456 catgcacctgtgtagacagatagatgagacttggtcaaatgtctatagaataa 1515
Db 1821 CGGGACCTGATTCGACAGATATGAAATGAACTTAAAGATATGTTAAGAAAGCA 1880
QY 1516 tgcgatggaaagcagccattatatacactgatacactaagaagaactctatata 1575
Db 1881 TGTATATATAGAGCTGACATTTATTAACGTTGACTACCAACAGAACAAATATCATT 1940
QY 1576 tggaaagggaaacacatcttaattaaatggtgcaacaaggaacctccggaagtc 1635
Db 1941 TGGAAAGAGAGATATCTCTATTAAACAAATGCGATGAGCTCCACCGGAAGCCCTTAT 2000
QY 1636 ctgaagtcatacacaacactagatgcatccaatgctcagggatagagaatgtaata 1695
Db 2001 CTGTAATCATATGAAATTAAGAGATACAGCTGCGCAGAGATTAGAGAAATGATA 2060
QY 1696 aaatcatcacaataacactagaagtagaagaacaatatacagaaagtgatatacca 1755
Db 2061 AAACACCCACTATAACCAATTTGAAGTTGAAGCTCATATATAGCAAGAACTTATGAT 2120
QY 1756 ttacaa 1761
Db 2121 ATACAA 2126

```

```

AC006279/c AC006279 245802 bp DNA HTG 12-AUG-2000
LOCUS Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***; 5 unordered pieces.
ACCESSION AC006279
VERSION AC006279.8 GI:9797723
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malarial parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 245802)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 245802)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT
On Aug 12, 2000 this sequence version replaced gi:8810451.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 147568: contig of 147568 bp in length
* 147569 147768: gap of unknown length
* 147769 217470: contig of 69702 bp in length
* 217471 217670: gap of unknown length
* 217671 219372: contig of 1702 bp in length
* 219373 219572: gap of unknown length
* 219573 221333: contig of 1761 bp in length
* 221334 221535: gap of unknown length
* 221536 245802: contig of 24269 bp in length.
FEATURES
source 1. 245802
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PRYAC614"
/clone="3D7"
BASE COUNT 99718 a 25978 c 21493 g 97813 t 800 others
ORIGIN
Query Match 11.1%; Score 196; DB 2; Length 245802;
Best Local Similarity 53.1%; Pred. No. 1.7e-19;
Matches 543; Conservative 0; Mismatches 455; Indels 24; Gaps 5;
QY 265 agaaatccacaagacaattttagaggatataaaaaagaaacaaaggtatatagca 324
Db 45336 AGAATAAACCCGAGAGAAATTAATTAATAAGAGATGGGAAGAAATTAATCA 45277
QY 325 ggaatcattggttctgaagcaagtcgaagtgatagagcaaaacataatgattcaaac 384
Db 45276 GGTATATTCATCATATGAGGTTCTTTATATATATA-----ACTTGAAGATTTAAT 45226
QY 385 ccaatggtcacaatttatataatgcatlgtcgcataatagttatagatgatact 444
Db 45225 CCTAATTTAATGTTGATACATATGCTCTTCTTAAGATGATTTATCCTATGACAGATA 45166
QY 445 agaccatttaatggaagcaaaagatccatataagaaacacggtcttagaataagaaac 504
Db 45165 GAGAGTATGTTGGTCTCCCTTTACTTTTCAAGTCATTAATAGGTTAAGAAATATATAGG 45106
QY 505 tatgtagatgcttaatgaataagcgtatcagaaagtagcgtcagatgaatatt 564
Db 45105 TTAACGATATTTTAAATGATGACATGATTTAAGAAAGCTGCTCGATATTTATATG 45046

```







Db	521	AACCCAGCGCAGCTTTGATGTCGTAATGATTAAGAAAAAGCGACTTGTGTAACAT	580
Qy	743	taattaactttttaagaaaaactatltccgaagaagaagttaattcaattctggttcac	802
Db	501	ACGTACCCAAACTCGCGCAATATACATGTGTGATTAACATGATCAGCATAGCCAGCCG	640
Qy	803	caaatgcgcattatcaatggtttccagagtgatgactt-----ctgttaagaatgaag	856
Db	641	CTAATGGCGCTCTGAGCTGCATCGGTTTCAACGACGCCGAAAAAATCTGTATGATGACG	700
Qy	857	aatctccatataaactaaattttgtcttgacaacatagaaacaaataagaattacata	916
Db	701	AAGCTCCGTAATAACAGCAAAATATT---CAACAACCGGACGGTTAAAGAAAGACTGTAC	757
Qy	917	ggcgagcgagatgtatcagcagaagaacttttaataattttaataacgaaagaga	976
Db	758	GGCGAGCCCAATGCGCATCTCGGGGTGGTGCATCTACTGATGAACACCTTGAAAGATA	817
Qy	977	aaatagatctgtattatlttaacacatacaatt--tagaactacaaatccagatataa	1033
Db	818	TGATTGACATGATGTTTGTGTGCAGAGTTCACTACACAGACTGACGATTTACCGTTA	877
Qy	1034	tgttagatagttacttatcccatlataatttggttttaaaaaataacatacatat	1093
Db	878	TGAAGAAATTATACGACTCTACGCATACATAGTATAAAAGTAGATAGTATGATATCA	937
Qy	1094	tagtttttcaatlgaaacatacagaagtgtgatttagtccgaanaataagaattatag	1153
Db	938	TGGGTTACACCTGATGTGTCCCGTCCACGCGGTTCAACCGGAAAGATATAATCTGTAA	997
Qy	1154	aattggttaggaaaaaacatacatgataaataaataatagtaggcagatggtatag	1213
Db	998	AATCATATCGCGCATTTCTGTAAAACTGAAAACAAACTGAATTAACCGCGAGATGTTG	1057
Qy	1214	ggaataggcatltaatt 1228	
Db	1058	GCCTGTGCTCTCTGCT 1072	
RESULT	8		
LOCUS	AC005504		
DEFINITION	Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS		
ACCESSION	AC005504	104992 bp	DNA
VERSION	AC005504.3		HTG 01-APR-1999
KEYWORDS	HTG; HTGS; PHASEL.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryotes: Alveolata: Apicomplexa: Haemosporida: Plasmodium.		
AUTHORS	1 (bases 1 to 104992) Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurd, O.B., Conway, A.B. and Davis, R.W.		
TITLE	Plasmodium falciparum 3D7 chromosome 12		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 104992)		
AUTHORS	Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-NOV-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Apr. 2, 1999 this sequence version replaced gt:4337172. * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 58642: contig of 58642 bp in length * 58643 58842: gap of unknown length * 58843 91011: contig of 32169 bp in length		

[illegible]







Query Match	Best Local Similarity	Matches	Conservative	Score	DB 3	Length	67970
549	44.9%	650	0	102.6	DB 3	67970	
9	44.9%	650	0	102.6	DB 3	67970	
8240	44.9%	650	0	102.6	DB 3	67970	
69	44.9%	650	0	102.6	DB 3	67970	
8300	44.9%	650	0	102.6	DB 3	67970	
129	44.9%	650	0	102.6	DB 3	67970	
8360	44.9%	650	0	102.6	DB 3	67970	
189	44.9%	650	0	102.6	DB 3	67970	
8420	44.9%	650	0	102.6	DB 3	67970	
249	44.9%	650	0	102.6	DB 3	67970	
8476	44.9%	650	0	102.6	DB 3	67970	
309	44.9%	650	0	102.6	DB 3	67970	
8536	44.9%	650	0	102.6	DB 3	67970	
369	44.9%	650	0	102.6	DB 3	67970	
8587	44.9%	650	0	102.6	DB 3	67970	
429	44.9%	650	0	102.6	DB 3	67970	
8644	44.9%	650	0	102.6	DB 3	67970	
489	44.9%	650	0	102.6	DB 3	67970	
8704	44.9%	650	0	102.6	DB 3	67970	
549	44.9%	650	0	102.6	DB 3	67970	



Db 209350 TAAATTAACAAAAAATAAANNNAATTTAATATATATCAAAAATATAT 209291  
Qy 311 aaggtatattagcagatctatgltcatggaacagtcagatgagcaaacata 370  
Db 209290 ATATAATATATCTTTTATTAATNCAATNTNTAATATTTTATTAATAAAAAA 209231  
Qy 371 tggattgtcacaaccatggtgctcaattatataatgcatctgtccgca--ttaata 427  
Db 209230 ATATTTTATATCATTAATAAAAAAACAACAATAATATTAATCTAAACCAACTATTA 209171  
Qy 428 ttttatgtatgltcatgagcattatggaagacaagttcctatgaagaacgcg 487  
Db 209170 TATTTATTTTACACCTAATAAATAATTTATTTATTAATAAATTTTATTAATAA 209111  
Qy 488 gcttagaataagaacctatgltcatgctccta--atgaattagacgtatcagaata 543  
Db 209110 TAAAAATATATACACNNATATTAATAATACAAAAAATTTTAAANAATTTTAAANA 209051  
Qy 544 gtacgtccagatgtaattatcttctatccttaggtgagagaacctatagatagata 603  
Db 209050 ACTAATATATATATTTTAAANNNAATAAATAATTAATAAANAATAATATATATTA 208991  
Qy 604 gaaaagaattgltcatgltcatggaataattggaacgtgtatgattgattgattgatt 663  
Db 208990 TTAATATNNNAATTTTATTAATAAAAAAACAATTTATTTAATAAATNAAATATCAAN 208931  
Qy 664 gttgtgatactgacgtggaacacacatggaagtttcaactaaatgaaatlaattt 723  
Db 208930 AATTAATAATATTTTAAANAATAAATAATTTNNNTATTAATTAATAAANAATNANT 208871  
Qy 724 tcaaatattatataatlaactaattatgltcatggaataa--acattccggaagaaagt 781  
Db 208870 CAAANATTAATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 208811  
Qy 782 taattcaattcgtgtcatgcaaatgctgattcatcagtggttcagaggtgattcctt 841  
Db 208810 AATTTTAAANNNNNCAATTAATAAAT----AATTNTAAATTAATAAATAAATAAATAA 208755  
Qy 842 tctgtaaagatgaagaatctccatataacactaaatttctgtcgaacaaatagaacaa 901  
Db 208754 AATATTAATAAATTAATAATATTAATATTAATTAATAAATAAATAAATAAATAA 208695  
Qy 902 ataaagattacatagagcgccgagtggtatcagcagaagaccttattatattt 961  
Db 208694 AATTAATAAATCTAANNCAANNNAATAAATAATTAATTAATAAATAAATAAATAA 208635  
Qy 962 atacagcaaggaagaatgattctgtattatcacaacatacacttgaagaactacaa 1021  
Db 208634 ATTAATAAATTTTACAAA--ACANAATTTTANATTAATTAATTAATTAATTAATTA 208576  
Qy 1022 atccagatataatgtagatagctactcactatataattgtgttlaaataataca 1081  
Db 208575 ATTAATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 208516  
Qy 1082 tcaaatcatataggttttcatcttagaacaatacagaagtggaattgattccggaata 1141  
Db 208515 TATTAATAAATAAATAAATAAANNNTNNAATTTAAACCAAAAAATTAATAAATAA 208456  
Qy 1142 aagaattattaggaattgtgtggaagaacaatacatgataaatacaaatatgaatgg 1201  
Db 208455 NCNTATTTAAATTCCAAAATTAANAATAATTAACACAAAAAATAAATAAATAA 208396  
Qy 1202 cagat 1206  
Db 208395 CATAT 208391

RESULT 12  
AC012492/C AC012492 183584 bp DNA PRI 20-MAY-2001  
DEFINITION Homo sapiens clone RP11-224B10, complete sequence.  
ACCESSION AC012492  
VERSION AC012492.9 GI:14151030

KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 183584)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 183584)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (28-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 183584)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On May 20, 2001 this sequence version replaced gi:13431121.  
Center project name: H.NH0224B10.  
FEATURES  
source  
BASE COUNT 53876 a 33878 c 34014 g 61816 t  
ORIGIN  
Query Match 5.3%; Score 93.8; DB 9; Length 183584;  
Best Local Similarity 44.4%; Pred. No. 6.2e-05;  
Matches 558; Conservative 0; Mismatches 687; Indels 12; Gaps 4;  
Qy 7 tttaaatatcaatatttcaattatagatcaccattgtattctgcaattccgaacc 66  
Db 16436 TTAATATTAATAATATTAATAATTAATAATTAATAATTAATAATTAATAA 16377  
Qy 67 ttgaaggaagaaataataataatcattcgtgggaataatcaggaagaaataataa 126  
Db 16376 TATTAATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAA 16317  
Qy 127 actcacaacggaataacatagatgcttccactcaatcaatcgaataatagtaatt 186  
Db 16316 TTAATATTAATAATATTAATAATTAATAATTAATAATTAATAATTAATAA 16257  
Qy 187 gtagaatacgatcttattgctggaagatgggtgaacttagaattacaaaataataa 246  
Db 16256 ATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 16197  
Qy 247 aactataataaataatgtagaataatcaccagaacattttagagagataaaaaagg 306  
Db 16196 TTAATATTAATAATCAATTAATTAATAATTAATAATTAATAATTAATAATTA 16137  
Qy 307 aaacaagttattagcagatactatggttctcgtggaacagtcagaggtatagacaa 366  
Db 16136 ATATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 16077  
Qy 367 catatgattgattcaaacccaatgg-----gtcaattatataatgcatctgtccga 421  
Db 16076 TATATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA 16017  
Qy 422 ttaatatgtatagatgltcatcagaccatttaattgaagaagaagattcctataagaa 481  
Db 16016 TATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 15957  
Qy 482 aacaagcttagaataagaacctatggtatgattgattgattgattgattgattgatt 541  
Db 15956 TATA-----TTAATAATGAATATATGTATTAATAATTAATAATTAATAATTA 15902  
Qy 542 aagtagctcagatgcaattatcttlatccttaggtggaagaacctatagatagata 601

Db	15901	ATATATATAGCATTTAAATATATCTATTTAAATATAGTATATATTAATATATATATTA	15842
Qy	602	tagaaaaagaatgtgattatgtagtaaaataatgtaagctgttaagatttgattag	661
Db	15841	AAATATATATTTAAATATGAAATATATATATATATATATATATATATATATATTA	15782
Qy	662	atggtgtagtattgactgtaggaacccatgaggaagtttaccacttaattgataaact	721
Db	15781	TATTA	15722
Qy	722	tttcaaatattatataatataatattactgtttaagaaaactatccggagaagaagc	781
Db	15721	TTAAATAATTA	15662
Qy	782	taatttcaattctggtgtcattcaaatgctgcatcattcatgctgttcaaggagttgac	841
Db	15661	ATTA	15603
Qy	842	tcgtgaagatgtagaagatcccatataacactaaattttgtctgaacaaatagaaca	901
Db	15602	TATTA	15543
Qy	902	ataaagaatttaccatagggcagcagcagtgctacacagcaagacttataattttta	961
Db	15542	TTAAATTA	15483
Qy	962	atacagcaaaaggagaaatagatcctgtatttcttcaaacatacaatttagaactaca	1021
Db	15482	ATTA	15423
Qy	1022	atccagaataaagtgtagatgttactatcccatattatttggtttcaataataca	1081
Db	15422	AACATTA	15363
Qy	1082	tcacatcatattaggttttccattcgacaatacagagggagattta-gtccggaat	1140
Db	15362	ATTA	15303
Qy	1141	aaagaattatgaattatgtaggaacaaacatacatgataaataacaaataaattag	1200
Db	15302	ATTA	15243
Qy	1201	gcagatggtatgaggtatgcatattatattgaagaagaacattaccacatgattca	1257
Db	15242	ATTA	15186
RESULT	13		
PPMAL3P5/c			
DEFINITION	PPMAL3P5	86827 bp	DNA
ACCESSION	AL0034556	AL008971	AL008972
VERSION	AL0034556.3	GI:7711065	
KEYWORDS	HTG; centromere; CTRP protein; initiation factor E4; Serine/Chreonine protein phosphatase.		
SOURCE	Plasmodium falciparum		
ORGANISM	Plasmodium falciparum		
REFERENCE	1 (bases 1 to 86827)		
AUTHORS	Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holtroyd, S., Hornsby, T., Horrocks, P., Jagsels, K., Jassal, B., Kyes, S., McLean, J., Mounle, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G.		
TITLE	The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum		
JOURNAL	Nature	400 (6744),	532-538 (1999)
MEDLINE	99376085		

[illegible]

	misc_feature	/gene="PFC0581w"	8642..8647	/note="potential splice donor sequence for exon 6 of PFC0581w"
	misc_feature	/gene="PFC0581w"	8704..8712	/note="potential splice acceptor sequence for exon 7 of PFC0581w"
	misc_feature	/gene="PFC0581w"	8913..8923	/note="potential splice donor sequence for exon 7 of PFC0581w"
	misc_feature	/gene="PFC0581w"	9007..9012	/note="potential splice acceptor sequence for exon 8 of PFC0581w"
	misc_feature	/gene="PFC0581w"	9112..9121	/note="potential splice donor sequence for exon 8 of PFC0581w"
	misc_feature	/gene="PFC0581w"	9259..9264	/note="potential splice acceptor sequence for exon 9 of PFC0581w"
	misc_feature	/gene="PFC0581w"	9361..9368	/note="potential splice donor sequence for exon 9 of PFC0581w"
	misc_feature	/gene="PFC0581w"		
	Query Match	5.3%; Score 93.6; DB 3;	Length 86827;	
	Best Local Similarity	45.1%; Pred. No. 7.3e-05;	Mismatches 478; Conservative 0; Mismatches 569; Indels 13; Gaps 3	
QY	225	tgaatcacaacaaataataataataaagaatgtagagaaatccagacaat	284	
DB	41295	TATAAATAATTAACACATCATATAAATCACTTATTAAATATATATATA	41236	
QY	285	tttagagggtcttaaaaaggaaacaaagtatttagcagatactatggtcatgaa	344	
DB	41235	TTTTAATTTTATTAATAATTAATTAATAATTAATAATTAATAATTAATTCATAAAA	41176	
QY	345	cagtcagggtgtaggcacaacatcgtatgccaaccatggcgtaacttata	404	
DB	41175	TATTTAAATATACATATATATTAATTTAAATATATATATTAAT--AATTAGGTATAT	41118	
QY	405	tatgcatttgctcgcataataatgltatatagatgatcagaccattaatggaaga	464	
DB	41117	ATATATATATTAATAATTAATAATTTTAATTTTAAATATTAATTTAAATAAATTCCA	41058	
QY	465	aagatgccattaaagaacacggcttagaataatgaacctatggtatgacttaaga	524	
DB	41057	TAAAATTAATTAATACATTAATTAATTAATAAATTTAAATAAATAATTAAT--TTAAATA	41002	
QY	525	aattagacgtacagaagaagtcgcagatgtaattatctttacacctagcgtgaga	584	
DB	41001	TTAACTATATATATATATTAATAAATAAATTAATTTATATTAATTTAGATTAAATTATA	40942	
QY	585	aacctatagatagatalagaaaaagaattgattatgtaataaatatcgtgaactigt	644	
DB	40941	TTATTAATAATTAATTAATAATTAATTTTATTAATTAATTAATTTATATATATAT	40882	
QY	645	taatgatcttgcttagatggtcgtatatactggaaccacatcgggaagtttacaa	704	
DB	40881	AAATTAATTAATAATTAATTAATTAATTAATTTAAATAAATAATGTTTAAATA	40822	
QY	705	ctaataagataaattttcaatlatatataatataatlaactcgttagaanaaac	764	
DB	40821	TTAAATTTTAATTAATTTTATTAATTAATAAATAATTAATTAATTAATTAATTAAT	40762	
QY	765	tattccgaagaagatcattcgaattcgttcattcaacaatgctgcatcacatgct	824	

```

Db      40761 TAAATTTAATTATTAATTGAAAATATAATGAATAGTAAAAATTAATTAATTAATTA 40702
Oy      825 ttccgaggttcacatccttcttgtaagaataaccatccatatcaactaaattttgc 884
Dd      40701 AATATATTTATTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40642
Oy      885 tgaacaatatagaacaataaagaatttacataggcgacggagtgttatccgacgaac 944
Dd      40641 ATATTTATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40582
Oy      945 tttaaatacatlittaalacagcacaggagaaaatagatcttgtattatccaacata 1004
Dd      40581 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40522
Oy      1005 caatttagaacatcaaatacccagataataatgltagatataigtacctatccccattaatt 1064
Dd      40521 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40462
Oy      1065 tggtttaaatlaatacacacacacatalataggtttlcataagacatacagaagtg 1124
Dd      40461 TATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40402
Oy      1125 attgatgccgaataaagaattttagaatttgtagaanaacaatacatgatataaaa 1184
Dd      40401 ATTAAAGTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40349
Oy      1185 tcaaaaataataataggcgacgtgatagsgatggcatttatatgaagaacaatt 1244
Dd      40348 AATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40289
Oy      1245 accaacitgataatcgcattgtagataattttctaccaat 1284
Dd      40288 ATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40249


RESULT_14
AC006281 LOCUS AC006281 199551 bp DNA HTG 12-AUG-2000
DEFINITION Plasmodium falciparum chromosome 12 clone 3D7, ** SEQUENCING IN PROGRESS **, 2 unordered pieces.
ACCESSION AC006281
VERSION AC006281.8 GI:9797738
KEYWORDS HMG; HTGS-PHASE1.
SOURCE malaria parasite P. falci-parum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 199551) Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T., Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12 Unpublished
JOURNAL 2 (bases 1 to 199551) Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
REFERENCE Direct Submission Submitted (05-JUN-1999) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT On Aug 12, 2000 this sequence version replaced gi:8810457.
** NOTE: This is a "working draft" sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* * * 1 41526: contig of 41526 bp in length
* * * 41726: gap of unknown length
* * * 41727 199551: contig of 157825 bp in length. Location/Qualifiers
FEATURES             location=Plasmodium falciparum"/organism="Plasmodium falciparum"
source

```

Query Match	5.3k:	Score	93.6:	DB 2:	Length	199551:		
Best Local Similarity	43.7k:	Pred.	No. 6.5e-05:					
Matches	569:	Conservative	0:	Mismatches	719:	Indels	14:	Gaps
QY 20	taattttaaataagatcaccatctgtatctgtcgaataatccagaaccttgaaagaaa	79						
Db 177411	TATGATTTACATATCATCTCTCTCTTCGATGCCGTCATTATTAAGCATTTTAAAAA	177470						
QY 80	ataataataataatcattcattgggaataacgagaaataaataaactcatcaacg	139						
Db 177471	AAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATAT	1775300						
QY 140	aataacagagctctttcacatccttaactcgaataagtaagtaatttgytaataagat	199						
Db 177531	AAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTA	1775500						
QY 200	cttatcgcgagatggygtatcactcgaattacaaaataaataaataaataa	259						
Db 177591	ATATATGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1776500						
QY 260	atgataagaatccca - gacaattttagagagatgaataaagaaagaaagat	317						
Db 177651	CTGAATTTATTTAAAAATTAACACTTAACCAATTAACCAATTAACCAATTA	1777100						
QY 318	tatagcagatcactatggtlcatggaacagtcgaagtgatagcgaataatgata	377						
Db 177711	ATCATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1777700						
QY 378	ttcaaacccaatggtatcattatataatgattcgtccgactaataatgata	437						
Db 177771	TCGTAAACATTTTACAAATTAATTTATTAATTAATTAATTAATTAATTAAT	1778300						
QY 438	tgatctgacacattatggaagaacaaagttccttaagaacaacgcttagata	497						
Db 177831	GTTAATTTCTTCTTTCATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1778900						
QY 498	tgaacccaatggtatgctttaaataagataagcgtatcagaaglacgcagat	557						
Db 177891	TAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1779490						
QY 558	aattatctttatcccttagtggaagaacctatagatagatagaaagaaatga	617						
Db 177950	AATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1780090						
QY 618	ttatggtgaataaataatgaaagctgtttaaagatttgatttagatgagatg	677						
Db 178010	TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1780690						
QY 678	ctggaagaacacatggaagtttcacactaaagaatlaatttcaatataat	737						
Db 178070	ATTATATCTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1781290						
QY 738	taaatlaacttactgtttaagaanaaactatcggagaagaagttlaattcaatt	797						
Db 178130	TTTTTTTTTATCTTTGGAAGAACAAATTAATTAATTAATTAATTAATTA	1781890						
QY 798	ttcatcaaaatgcatcatcaatgcgtttcagaagttgcatcttctgtaaga	857						
Db 178190	TAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1782490						
QY 858	atccacataaactaaatttctgtgacaataagaaacaaataaagaattacag	917						
Db 178250	ATATTACTTTTATATATATATATATATCTTTTATACATTAATTAATTAAT	1783090						
QY 918	ggcagcagcgatgcatcagcagaacatttataatattttaaatacagcaagg	977						

Db 178310 ACTACTTTTCATTATGATAGAGAAAGAGCAATATCTCTCTTTATTTGTTGAATTAAATC 1783369  
OY 978 aatagatcttgatct-----tatcaaacatacaatttagaacaataatcca 1026  
Db 178370 ATTTGTTGATGTTAAATGCTAAATAATATCAACATTAATATATATATATATATAA 178429  
OY 1027 gataataagtagatagctacttaaccattatatttggtttaaaataacatacacaca 1086  
Db 178430 TAT 178489  
OY 1087 atcatatagtttttcaatcagaacaacagagcttgattagtcgccgaataaagaa 1146  
Db 178490 ATTGACCTGGAATTAATTTCTTATTTATTTATATATATATATATATATATATATAT 178549  
OY 1147 ttattagatcttgtaggaagaacaacatacatagataaaatacaataatagagcagat 1206  
Db 178550 CTTCCTTTTATTTATTAACAAAAAAGAAAGAAAAAATAATTAATAATAATAATAATA 178609  
OY 1207 gataagagatagcattatattatgaagaacaacataccactgagatcagatga 1266  
Db 178610 AATAAATAAATAAT 178669  
OY 1267 gatattttcttaacaatatattggaacaacttaaatcctgaa 1308  
Db 178670 TATATATTTAGTACCAATTTGAAAGAACTATTAAATTAATA 178711

RESULT 15  
AF396436/c 47577 bp DNA INV 25-JUL-2001  
LOCUS Tetrahymena thermophila strain SB210 mitochondrion, complete genome.  
DEFINITION AF396436  
ACCESSION AF396436  
VERSION AF396436.1 GI:15011465  
KEYWORDS  
SOURCE Tetrahymena thermophila.  
ORGANISM Mitochondrion Tetrahymena thermophila  
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;  
Hymenostomata; Tetrahymenina; Tetrahymena.  
REFERENCE 1 (bases 1 to 47577)  
AUTHORS Brunk,C.F., Tran,A.B., Lee,L.C. and Li,J.  
TITLE Complete Sequence of the Mitochondrial Genome of Tetrahymena thermophila and Comparison With the Mitochondrial Genome of Tetrahymena Pyriformis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 47577)  
AUTHORS Brunk,C.F., Tran,A.B., Lee,L.C. and Li,J.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2001) Organismic Biology, Ecology and Evolution, University of California, Los Angeles, 621 Charles E. Young Drive South, Los Angeles, CA 90095, USA  
FEATURES  
source 1..47577  
/organism="Tetrahymena thermophila"  
/organelle="mitochondrion"  
/strain="SB210"  
/db\_xref="taxon:5911"  
repeat\_region complement(1..53)  
/note="telomere repeat"  
trna complement(65..145)  
/gene="trny"  
/product="tRNA-Tyr"  
/anticodon="(pos:110..112,aa:Tyr)  
gene complement(65..145)  
/gene="trny"  
misc\_feature complement(161..2844)  
/note="copy number 1; includes rnl\_a, trnl and rnl\_b"  
rna complement(161..438)  
/gene="rnl\_a\_1"  
/product="large subunit ribosomal RNA"  
gene complement(161..438)  
/gene="rnl\_a\_1"

trna complement(448..528)  
/gene="trnl\_1"  
/product="tRNA-Leu"  
/anticodon="(pos:492..494,aa:Leu)  
gene complement(448..528)  
/gene="trnl\_1"  
rna complement(531..2844)  
/gene="rnl\_b\_1"  
/product="large subunit ribosomal RNA"  
gene complement(531..2844)  
/gene="rnl\_b\_1"  
complement(3023..3325)  
/gene="ymf57"  
complement(3023..3325)  
/gene="ymf57"  
CDS  
/note="CPGN"  
/codon\_start=1  
/transl\_table=4  
/product="ymf57"  
/protein\_id="AAK7561.1"  
/db\_xref="GI:15011466"  
/translation="MLKNLIKRFKFRFVQSGPYDIFPKFSEMFIRNIFYSIF  
GKPFMEYLKRTIDSFIRNNRNFNINVESKYPQLITLLILYLFITIFILPII"  
gene complement(3335..4675)  
/gene="ymf66"  
complement(3335..4675)  
/gene="ymf66"  
CDS  
/note="CPGN"  
/codon\_start=1  
/transl\_table=4  
/product="ymf66"  
/protein\_id="AAK7562.1"  
/db\_xref="GI:15011467"  
/translation="MGRENVLPHVNDVYEDFVFTPIPYQPESTFKSVPLKSDILGG  
VEMVYTTSSVLAYDKYKIMSGVSNLDESDMFNFQWALSTSVQFLYANITD  
RILSVLPQNTPTNDWFRMLHSHKETALIMLYHPELSMRTINGLQFRTYFGGLIEFV  
YEDKSNPDMCIIIVHTLMIHLILFLFTGFVILTSFGNPTBENTIDSDLAASG  
VEAEKRTISIDVILGLVFAIVGVFVHGMVSMLSHAYLLSCYSIIIMFLFLLG  
MPLLALDITGIFELAYLKGAGKISSVAAEMQWACLVFIRILAQWIRVILMYTF  
ISLSHVSPEDITNSALISGENSDSMENLNFQMTYIILVLPKFTYMYEILHT  
FVYVCSQFAFPAIVFWLPIFLYPIIIEKHDFSKREBKRLKLEMLNKN"  
gene complement(4937..6154)  
/gene="ymf76"  
complement(4937..6154)  
/gene="ymf76"  
/note="CPGN"  
/codon\_start=1  
/transl\_table=4  
/product="ymf76"  
/protein\_id="AAK7563.1"  
/db\_xref="GI:15011468"  
/translation="MPLVKEFIKKLRLKRIKKFYKLINYNFNLNNFYHKKPNFEL  
LVNNTNPELKITLFIYKINLISKITSNKLPLNNKRIRRRKKFKYKYSKDILKIT  
NILKSKTKIKNLFPRIKFSVLARKROSKRIPMLAVKSRISRRRFFKKLKKKIAY  
FTQWRKLLKFKKRTIINLINDFTIRNRKFRHLRLDIRKTIIRYLNRRNITKGF  
KRIKINFEKIRKHKISIKRIVHPELISLNIAIKRYATFLRNANLFTKSGF  
IFLNGHOELNPFKAYAGDIIELPESKFLKLRKMKKRNMSMKYKRYKWWRLKKN  
VDEQRRLRISRESENTLNFKTKLTKLQYDYRTLSCVAVLDTPNKRDLTYLNKLPI  
IYLKLFNKTIIS"  
gene complement(6189..7019)  
/gene="rps13"  
complement(6189..7019)  
/gene="rps13"  
CDS  
/codon\_start=4  
/transl\_table=4  
/product="ribosomal protein S13"  
/protein\_id="AAK7564.1"  
/db\_xref="GI:15011469"  
/translation="MKEFQLNGYIFTDTEINTKKKIFGMNDYTYKIIYNREKFY  
SENLSYKDEWKFKRSITSPFKRSILKNMDNIFLIDPLANTYRGLRHSKGLPYR  
GORTWNASVRSNLIARKELEIARKIYGMPPESITTYIADQIYIKLQMKRE  
WLOAKSRKLKMQNHAKFKIDINSGYIDGPPKKKEITRKKKAQAKNTPFLUGRE  
PGFTLFLYLRPNNSLSAKERKIKIVLNEENKVILOKKRRVVLKKRPEKKKNSWE



gene  
complement(7023. .7478)  
/gene="tps3"  
complement(7023. .7478)  
/gene="tps3"  
/codon\_start=1  
/product="ribosomal protein S3"  
/protein\_id="AAK7565.1"  
/db\_xref="GI:15011470"  
/translation="MGLSKSPMLNKSIGSIWMYHNIWDSIKLYKKYSLSFLENEVINH  
FLNEMYYVICIKKIRPTDPRKLGFGRGNSININKIKSMNMRHFLGKILFLKYQGWV  
LYLVNYSRRKKILINYSFKAFKKIKASRQGVTSVYYKMDKIKRKF"  
complement(7483. .7779)  
/gene="tps19"  
complement(7483. .7779)  
/gene="tps19"  
/codon\_start=1  
/transl\_table=4  
/product="ribosomal protein S19"  
/protein\_id="AAK7566.1"  
/db\_xref="GI:15011471"  
/translation="WGSWMRLKYISKSIIKISIGIKIKKKISVINKNSSTIPS  
SLTDWTFPIHNGYKRELIKIDYHGFGEFIFILRRPNIFPKSKNKSFRF"  
complement(7784. .8572)  
/gene="rpl2"  
complement(7784. .8572)  
/gene="rpl2"  
/codon\_start=1  
/transl\_table=4  
/product="ribosomal protein L2"  
/protein\_id="AAK7567.1"  
/db\_xref="GI:15011472"  
/translation="MELKNYKRPITLRFKRDIKRMPFIKKNNIFETIKNKSQSGST  
KGQILVIRRGAKTYIKKHLISIVNRNQLNLNLIYSLNTINKNSCFGLIKRNSCLSY  
IKIAGTFLIGNTKSLDKPLNFSFNRILOSFTILMLAPKHSIFSLISNKKNSY  
AKSLAGTFLISIIKIKELIMFLIKLPTGEEKYFGSDSKATIGRNSNPLNKFYVIGKAI  
NNKNCRRSSVRGVANPVDPHRGHRTKNQPEMSPGWTKKK"  
complement(8576. .9049)  
/gene="ymf74"  
complement(8576. .9049)  
/gene="ymf74"  
/note="CPGN"  
/codon\_start=1  
/transl\_table=4  
/product="ymf74"  
/protein\_id="AAK7568.1"  
/db\_xref="GI:15011473"  
/translation="MKFRERFLIKKKNFKKKETILKIYISIKNLINFKYIKIKLN  
NFKIKSTILNNNYFNIITNGLDKYDPTQNEELWTLISIKYKKNLILISNNQD  
IKPKFSTIIDKNKYNSLIDCNLEDFEFTIILITNSLILEFYNIILINIKIN"  
complement(9062. .9550)  
/gene="nadi10"  
complement(9062. .9550)  
/gene="nadi10"  
/EC\_number="1.6.5.3"  
/codon\_start=1  
/transl\_table=4  
/product="NADH dehydrogenase subunit 10"  
/protein\_id="AAK7569.1"  
/db\_xref="GI:15011474"  
/translation="MRVDYIVTPINAVVSVMARQGSFMLEFGIACAVEMMHAVISY  
DEDFRGLVFERRAPROADRIILVAGLTITNMARGLRIRYEQTLDPKWLVSMSGCAAGGY  
YHYSVVRGCRNRIIPVDYVPGCPRAFAALMFGVLDLQKSLYQINDEGKKYKKRL  
TY"  
complement(9567. .9968)  
/gene="rps12"  
complement(9567. .9968)  
/gene="rps12"  
/codon\_start=1  
/transl\_table=4  
/product="ribosomal protein S12"  
/protein\_id="AAK7570.1"

gene	/db_xref="GI:15011475"	
CDS	/translation="MGSCIKKKKGNLHKRLKRAILIGINPOKKGVKKARIVTPRKPN SARRPAKVIILVNRKRLTAIPGIGHIIRHRSVLVGGCCRIDLPGRVYTCVGVDF SKVLRFRRSRYGIPKDEKKKKLRFRY"	
	complement(10019..10555)	
	/gene="nad2"	
	complement(10019..10555)	
	/gene="nad2"	
	/EC_number="1.6.5.3"	
	/codon_start=1	
	/transl_table=4	
Query Match	5.2%	Score 92; DB 3; Length 47577;
Best Local Similarity	44.5%	Pred. No. 0.00013;
Matches 679; Conservative 0; Mismatches 825; Indels 22; Gaps 7;		
QY	6	ttttaaatatcaacatttttaattatagatgcacatcttgattcttcgcaaatccagAAC 65
DB	26633	TTTTAAATTTTATTTTCATTTATTTATTTATTTAAACAAATTAATTTTTTTTAACTTTATTTATA 26574
QY	66	cttgaagaagaaanaataataaattccattcttggaataaatacggaaataaanaaA 125
DB	26573	CGGTATATATTTATTTATTAATAATTTTATGAGATTTATGTACCACTGTATTTATATGAGATTT 26514
QY	126	aactcaacaaacggaataacatgagctctttcacaccttaactaaatcgaaatagAAC 185
DB	26513	TATTAATGCTAATTTAAATTTATATATGCAATCTTAAATATATAGAAAAAAAATTTATTTGT 26454
QY	166	tgtA-----gaataggtcttatgtcgagagtggtgigttaactcagaattacaaA-aa 239
DB	26453	ATTAAATTCAAATTAATGAGATTTTCATTTATGATGCTGATTTATGATATTAATAATTAATAA 26394
QY	240	taataaanaataaataaataaatacagatagaaacacacaaagacaatcttagagagAAC 299
DB	26393	TGATATTAATAAATTTATCGAAAAATTTTAATCTGCAATTAATAAAAAAAAATTTAAAAAGA 26334
QY	300	aaaaaggaacaaaggatlatatagcsgagactactggttcatlagnaacagcagagtgaAC 359
DB	26333	AAAAATTTAATACATTAATATTTATTTATTTATTTGAAATTTTGCAATCTTATTAATAT 26274
QY	360	agcaaacacatgatgtgttccaacccaatgctgcacattatataatgatcattgcgcG 419
DB	26273	TAAATATGTTTTAAATTTTATTTATTTATTTATTTAATTTAATTTAGTATCTATTAATAT 26214
QY	420	catlaatagttatatgtactcagacacatlaaTgaagaacaaagatccctaaAC 479
DB	26213	ATATATTAATAAATTTTACAAATTAATTAATTTAGTAAATCTTGAATTTAGTTCAATACAA 26154
QY	480	aaacacagcgcttagaataTgaacctatgtaTgactgcttaatgaattagacgtacAG 539
DB	26153	AAAAATTTATTTCAAAATTTATTTATTTATTTATTTAATTTAAGATTTGAAAAATTTAAAAACCTT 26094
QY	540	aaagtaagT-ccagatgtaattatcttcttaaccttaggggggaacacacatagatAG 598
DB	26093	ATATGTAAATTAATTAATAAATTAATTTATCTACAAATTTTAACTTTCTTATATTTAAAGA 26034
QY	599	ataTgaagaagaatTgatitTgtaTgaataataTgaactTgtaTgaatttgatAC 658
DB	26033	AAAAAAAACGAAAAAATTTTGAAGCAATATTAATCTTAACTATTTTATTTTAAATTTT 25974
QY	659	tagatggtTtagataTgactTggaacacacatTggaagtTttacaacttaaatgaatA 718
DB	25973	TAAAAATTTTAAATATTTATTTATTTTAAATTTTAAATTTTAAATTTAAGATGATTTT 25914
QY	719	atttccaatatatact-----aaatlaaTcaactTgtTgaagaaaaaactatccgGA 773
DB	25913	ATTAAATTAATGATTAATTTTATTTAAATTTAACTTATTTTATTAATAAAGATTTCTGGA 25854
QY	774	agaagaatTaatTcaatt-----cgctTcatcaaatTcgtgaatcatcatcgtgcttc 827
DB	25853	AATTAATAATTAACCTTAATTTTATTTATTAATATTTTATTAATTTATTAATAATTAATATAT 25794
QY	828	aggagttcacatcttctctTaaagatTgaagaactctccatataacactaaatttttctcGA 887

```
Db 25793 ATATATTTTACATATATATACAAAATTAAATATATATATCTAATTAATATATAAAAAATTA 25734
QY 888 acaaatagaaacaataaagaattacataaggcgacgagtgtatcagcaggaactt 947
Db 25733 TAATATATATCTATATATATATATATATATATATATATATATATATATATATATATATAT 25674
QY 948 tattaataattttaaatacagcaagaagaataatagatctgtatttalttcaacatacaa 1007
Db 25673 AAGTAAATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25614
QY 1008 tttagaactcacaatccagatataat--gtagatatgtaactatuccatltatat 1064
Db 25613 ATTAATAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 25554
QY 1065 tggctttaaataataacatacacaatcataltaggttttcatlagaacatacagaagtg 1124
Db 25553 AAAAATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 25494
QY 1125 atttagtcccgaaataaagaataatttagaatttggtaggaaanaacaatalacatgataa 1184
Db 25493 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 25434
QY 1185 tcaaaataataatagagcgagatggtlagaagatalgcatttatttgaagaacaact 1244
Db 25433 TAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 25375
QY 1245 accaactggatcaltcgatgtagatatlttcttacaataatttggaaacattttaaacc 1304
Db 25374 GTAAAAATAGAAAAATATATATATATATATATATATATATATATATATATATATATAT 25315
QY 1305 tgaagtagcaaacctcacaagaacttactataactgaaacccctgaagaactgtagacaat 1364
Db 25314 TATTAGAAAAATATATATATATATATATATATATATATATATATATATATATATATAT 25255
QY 1365 agatgaatatgltccaggactcgltatccacaatagagatatatacaacaacaatga 1424
Db 25254 TAAATTAATAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 25195
QY 1425 tgcataatggaaactagatcttataatcattcatgcaactgctgtagagacagatatgact 1484
Db 25194 ATTTATACATTAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAAT 25135
QY 1485 ggacttggcacaagtagtctatgaa 1510
Db 25134 TGAATATATCCAAATGTTATTAATAA 25109
```

Search completed: March 21, 2002, 15:52:35  
Job time: 5086 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:29:29 ; Search time 174.24 Seconds

(without alignments)  
8679.532 Million cell updates/sec

Title: US-09-579-383-2

Perfect score: 1764

Sequence: 1 atgaatttaataatcaat.....tgattaccattacaataa 1764

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 930621 segs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database :  
1: /SID52/gcgdata/geneseq/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	1764	22	AC89669
2	1764	100.0	2500	22	AC89670
3	219	12.4	936	22	AA8252
4	219	12.4	936	22	AA8252
5	219	12.4	936	22	AA8252
6	219	12.4	936	22	AA8252
7	219	12.4	936	22	AA8252
8	219	12.4	936	22	AA8252
9	217.2	12.3	936	22	AA8252
10	217.2	12.3	936	22	AA8252
11	217.2	12.3	936	22	AA8252

12	217.2	12.3	936	22	AA8259	Oligonucleotide D2
13	217.2	12.3	936	22	AA8252	Oligonucleotide D2
14	217.2	12.3	936	22	AA8255	Oligonucleotide D1
15	104.2	5.9	1137	22	AA8252	P. falciparum chit
16	73.2	4.1	244	22	AA8238	Oligonucleotide D1
17	71	4.0	244	22	AA8238	Oligonucleotide D1
18	66.4	3.8	11922	21	AA70187	Plasmodium falcipa
19	63.8	3.6	7458	21	AA70106	Plasmodium falcipa
20	63.4	3.6	3101	11	AA02047	Sequence encoding
21	62.6	3.5	6621	21	AA70188	Plasmodium falcipa
22	62.2	3.5	19124	18	AA72882	Plasmodium var-7 g
23	62.2	3.5	19124	21	AA72827	Plasmodium var-7 p
24	61.2	3.5	3642	21	AA70180	Plasmodium falcipa
25	61	3.5	3399	17	AA05868	Chicken leucocytos
26	59.6	3.4	2503	15	AA053480	PNP30 xylanase CD
27	59.6	3.4	6042	21	AA70189	Plasmodium falcipa
28	59.6	3.4	9789	17	AA71852	CDNA encoding Plas
29	58.6	3.3	4838	20	AA70891	Plasmodium var-esp-1a
30	58.4	3.3	5940	21	AA70105	Plasmodium falcipa
31	58.2	3.3	5145	21	AA70209	Plasmodium falcipa
32	58	3.3	665	21	AA70196	Human apolipoprote
33	58	3.3	1671	13	AA024134	50 kD subunit of S
34	57.8	3.3	4197	21	AA70227	Plasmodium falcipa
35	57.8	3.3	6033	21	AA70152	Plasmodium falcipa
36	57.8	3.3	6644	20	AA33181	Base sequence of t
37	57.8	3.3	7372	20	AA33182	Base sequence of t
38	57.8	3.3	7797	20	AA33180	Cowpox virus bsr f
39	57.8	3.3	7996	20	AA33184	Base sequence of t
40	57.6	3.3	765	21	AA059300	Human secreted pro
41	57	3.2	612	22	AA71471	Human cervical can
42	57	3.2	660	15	AA066819	AMEPV thymidine-kl
43	57	3.2	1511	15	AA028302	AMEPV tk DNA. Ams
44	57	3.2	1511	15	AA066798	AMEPV thymidine-kl
45	57	3.2	1511	19	AA714508	AMEPV entomopoxvir

#### ALIGNMENTS

RESULT 1  
AAC89669 ID AC89669 standard; DNA: 1764 BP.  
AC AC89669;  
DT 13-MAR-2001 (first entry)  
XX  
DE P. gallinaceum chitinase coding sequence SPQ ID NO: 2.  
KW Malaria; mosquito; chitinase; fungal disease; parasitic disease;  
KW veterinary disease; arthropod pest; ds.  
XX  
OS Plasmodium gallinaceum.  
XX  
PN W0200073488-A1.  
XX  
PD 07-DEC-2000.  
XX  
PF 26-MAY-2000; 2000MO-US14536.  
XX  
PR 28-MAY-1999; 99US-0136508.  
PR 03-FEB-2000; 2000US-0180051.  
XX  
XX (TEXA) UNIV TEXAS SYSTEM.  
PI Vinetx JM;  
XX WPI; 2001-061553/07.  
XX P-PSDB; AAB49986.  
XX  
XX New nucleic acid encoding a Plasmodium species chitinase is useful for  
XX preventing transmission of malaria by mosquito feeding on subject that  
XX may harbor Plasmodium species organisms

XX Claim 4; Page 123-124; 137pp; English.

PS The present invention provides the protein and coding sequences of the  
CC Plasmodium falciparum and P. gallinaceum chitinase enzymes. These  
CC organisms are the cause of malaria in humans. The sequences are useful  
CC in the prevention and treatment of malaria, fungal diseases, parasitic  
CC diseases and veterinary diseases, in preventing the transmission of  
CC malaria and in the control of arthropod pests in agriculture.

XX Sequence 1764 BP; 717 A; 227 C; 294 G; 526 T; 0 other;

Query Match 100.0%; Score 1764; DB 22; Length 1764;  
Best Local Similarity 100.0%; Pred. No. 2,1e-287;  
Matches 1764; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 atgattttaaatcaatatttttaattatagatccatctgtatctcgcaattcc 60
DB 1 atgattttaaatcaatatttttaattatagatccatctgtatctcgcaattcc 60
QY 61 agaacttgaaagaaaaaataataataattcattggaataatacgggaaaaataa 120
DB 61 agaacttgaaagaaaaaataataataattcattggaataatacgggaaaaataa 120
QY 121 aataaactcaacaacggaataatcatgagtcctttccacatcthaatcgaaataagt 180
DB 121 aataaactcaacaacggaataatcatgagtcctttccacatcthaatcgaaataagt 180
QY 181 aatttttagaataatgatacttatttgcggaatgagtggttaactcgaataataa 240
DB 181 aatttttagaataatgatacttatttgcggaatgagtggttaactcgaataataa 240
QY 241 aataaataataataaataatgataaataacacacaagaacattttgagagratataa 300
DB 241 aataaataataataaataatgataaataacacacaagaacattttgagagratataa 300
QY 301 aaaaagaaacaggtattatagcagagatactggttcaatggaacagtgataga 360
DB 301 aaaaagaaacaggtattatagcagagatactggttcaatggaacagtgataga 360
QY 361 gcaaaatatgatttgatttcaaaccaatggtgtcaatttatattgattgtctgc 420
DB 361 gcaaaatatgatttgatttcaaaccaatggtgtcaatttatattgattgtctgc 420
QY 421 attaataatgataatgatactcagaacatttaataagaaacaaagaattccttaaga 480
DB 421 attaataatgataatgatactcagaacatttaataagaaacaaagaattccttaaga 480
QY 481 aaacacggttagaataatgaaacctatgatatgattgaaatagaagatcaga 540
DB 481 aaacacggttagaataatgaaacctatgatatgattgaaatagaagatcaga 540
QY 541 aaagtcggtccagatgaatatttcttcttccctttagtggagaacactatagatag 600
DB 541 aaagtcggtccagatgaatatttcttcttccctttagtggagaacactatagatag 600
QY 601 atagaaaaaagaattgattatgtaataaataattgaaagctgttgaatgattgattta 660
DB 601 atagaaaaaagaattgattatgtaataaataattgaaagctgttgaatgattgattta 660
QY 661 gatggtgtagataatgacttggaaacacatggaaggttttaacacttaaatgaat 720
DB 661 gatggtgtagataatgacttggaaacacatggaaggttttaacacttaaatgaat 720
QY 721 ttctcaaatatataatataatgaactgttaagaanaactcgcgaagaagaag 780
DB 721 ttctcaaatatataatataatgaactgttaagaanaactcgcgaagaagaagaag 780
QY 781 ttaattcaattctggttcaataaagtcgtcatatcagcgttttaggaagtgtcatc 840
DB 781 ttaattcaattctggttcaataaagtcgtcatatcagcgttttaggaagtgtcatc 840
```

```
QY 841 ttctgtaagaatgaagaatctccatataacataatttttgtctgacaataagaaca 900
DB 841 ttctgtaagaatgaagaatctccatataacataatttttgtctgacaataagaaca 900
QY 901 aataaagaattacatagggcagcagcgatgttatcagcaggaactttatataattt 960
DB 901 aataaagaattacatagggcagcagcgatgttatcagcaggaactttatataattt 960
QY 961 aatacagcaaggaagaataatagatctgttatcttaataataataattagaactaca 1020
DB 961 aatacagcaaggaagaataatagatctgttatcttaataataataattagaactaca 1020
QY 1021 aatccagataataatgtagatagatgtaactatcccatattatatttggtttaaaatatac 1080
DB 1021 aatccagataataatgtagatagatgtaactatcccatattatatttggtttaaaatatac 1080
QY 1081 atcaaatcatatagatgttttttcaattgaaataacacgaggtgatttaaccgaaat 1140
DB 1081 atcaaatcatatagatgttttttcaattgaaataacacgaggtgatttaaccgaaat 1140
QY 1141 aaagaattattagattgtagaagaaacaatacatgataataaatacaaatatagag 1200
DB 1141 aaagaattattagattgtagaagaaacaatacatgataataaatacaaatatagag 1200
QY 1201 gcaagatgtagaggaatagggatatttatttgaagaacaattaccactggatcttc 1260
DB 1201 gcaagatgtagaggaatagggatatttatttgaagaacaattaccactggatcttc 1260
QY 1261 gatgtagatattttcttacaataattttggaataatcctctgaaagataaactcca 1320
DB 1261 gatgtagatattttcttacaataattttggaataatcctctgaaagataaactcca 1320
QY 1321 aagaacctactataactgaaaccccgaaagactgtgcacaataagatagtctcca 1380
DB 1321 aagaacctactataactgaaaccccgaaagactgtgcacaataagatagtctcca 1380
QY 1381 ggaactcgttattccaacatagggatatttaacaacaacatgatactgataagaact 1440
DB 1381 ggaactcgttattccaacatagggatatttatacaacaacatgatactgataagaact 1440
QY 1441 agactctatttaattcctatgacccctggtgtgacagatgaatggagactgtgcaaga 1500
DB 1441 agactctatttaattcctatgacccctggtgtgacagatgaatggagactgtgcaaga 1500
QY 1501 tgcataagaaaataatcgaatggaagcagcccatattataaacctgactataaga 1560
DB 1501 tgcataagaaaataatcgaatggaagcagcccatattataaacctgactataaga 1560
QY 1561 agctctattatataatgaaagggaacacatatttaaatgaatggtggacaagaacct 1620
DB 1561 agctctattatataatgaaagggaacacatatttaaatgaatggtggacaagaacct 1620
QY 1621 ccggaagtgtaagcagcactaggtatatacaaaaactgatacccaatggtccagagata 1680
DB 1621 ccggaagtgtaagcagcactaggtatatacaaaaactgatacccaatggtccagagata 1680
QY 1681 gaaagatggaataaataatccacataaacccctagaagtagaaggaacataatgacaa 1740
DB 1681 gaaagatggaataaataatccacataaacccctagaagtagaaggaacataatgacaa 1740
QY 1741 gaagtgtattaccattacaataa 1764
DB 1741 gaagtgtattaccattacaataa 1764
```

RESULT 2  
AAC89670  
ID AAC89670 standard; DNA; 2500 BP.  
XX  
AC AAC89670;  
XX  
DT 13-MAR-2001 (first entry)  
XX

DE	P. gallinaceum chitinase coding sequence SEQ ID NO: 5.
XX	
KW	Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KM	veterinary disease; arthropod pest; ds.
XX	
OS	Plasmodium gallinaceum.
XX	
PN	WO20073488-A1.
XX	
PD	07-DEC-2000.
XX	
PE	26-MAY-2000; 2000MO-US14536.
XX	
PR	28-MAY-1999; 99US-0136508.
XX	
PA	03-FEB-2000; 2000US-0180051.
XX	
PI	(TEXA ) UNIV TEXAS SYSTEM.
DR	Vinetz JM;
XX	
XX	WPI: 2001-061553/07.
PT	New nucleic acid encoding a Plasmodium species chitinase is useful for
XX	preventing transmission of malaria by mosquito feeding on subject that
XX	may harbor Plasmodium species organisms -
PS	Disclosure: Page 127-128; 137pp; English.
CC	
XX	The present invention provides the protein and coding sequences of the
CC	Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC	organisms are the cause of malaria in humans. The sequences are useful
CC	in the prevention and treatment of malaria, fungal diseases, parasitic
CC	diseases and veterinary diseases, in preventing the transmission of
CC	malaria and in the control of arthropod pests in agriculture.
XX	
SQ	Sequence 2500 BP; 993 A; 283 C; 365 G; 858 T; 1 other:
Query Match	100.0%; Score 1764; DB 22; Length 2500;
Best Local Similarity	100.0%; Pred. No. 2,le-287;
Matches 1764; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 atgaattttaaatacacaatcttttaattatagratcacattgttatctgcgaattcc 60
DB	276 atgaattttaaatacacaatcttttaattatagratcacattgttatctgcgaattcc 335
OY	61 agaacttcgaaaggaaaataataataaatcatcttgtagaataacgggaaataaa 120
DB	336 agaacttcgaaaggaaaataataataaatcatcttgtagaataacgggaaataaa 395
OY	121 aataaacatcatcaaacgggaaataacatagatgcttttcacatcttaaatcgatalagt 180
DB	396 aataaacatcatcaaacgggaaataacatagatgcttttcacatcttaaatcgatalagt 455
OY	181 aatttgtagaatagatgatcttatcttcgagagatgggtgaactctgatatcaaaaaat 240
DB	456 aatttgtagaatagatgatcttatcttcgagagatgggtgaactctgatatcaaaaaat 515
OY	241 aataaaatatataaanaaatgatagaaataccaagaagcaaatcttgaagaagataaa 300
DB	516 aataaaatatataaanaaatgatagaaataccaagaagcaaatcttgaagaagataaa 575
OY	301 aaaaggaacaagatcatatagacagagatactatggttcagtggaacagtcgaagtgataga 360
DB	576 aaaaggaacaagatcatatagacagagatactatggttcagtggaacagtcgaagtgataga 635
OY	361 ggaagaacatagatgatccaaccgaatggtgcctaatttatatgatcttcctgc 420
DB	636 ggaagaacatagatgatccaaccgaatggtgcctaatttatatgatcttcctgc 695
OY	421 attaatatgtatataatgatctctagaccatttaatggaagaacaagaattccatataaga 480
DB	696 attaatatgtatataatgatctctagaccatttaatggaagaacaagaattccatataaga 755

QY	441	aaacagcgcttagaataatgaacctaagatgtaatgcttaatgaaatagacgatacga	540
DB	756	aaacacgctcttagaataatgaacctaagatgtaatgcttaatgaaatagacgatacga	815
QY	541	aaagaagctccagatgtaataatctcttctacgttagggggaacccataatgataat	600
DB	816	aaagtaagctccagatgtaataatctcttctacgttagggggaacccataatgataat	875
QY	601	atagaanaaagaatgttaatgtgataaataatgaacgtgttaatgtattgtattca	660
DB	876	atagaanaaagaatgttaatgtgataaataatgaacgtgttaatgtattgtattca	935
QY	661	gatggtgtgaatattgctctggaacccaatagggaagttttacaacttaatgataat	720
DB	936	gatggtgtgaatattgctctggaacccaatagggaagttttacaacttaatgataat	995
QY	721	ttttcaaatatataataatgaatgaactgtttaagaanaacattcccggaagaag	780
DB	996	ttttcaaatatataataatgaatgaactgtttaagaanaacattcccggaagaag	1055
QY	761	tttaattcaattctggttccaacaaatgcgtcatatcaatgcgtttcagaagtgcact	840
DB	1056	tttaattcaattctggttccaacaaatgcgtcatatcaatgcgtttcagaagtgcact	1115
QY	841	ttctgttaagaatgtgaagaattccataaacaataatttttctggaacaaatagaaca	900
DB	1116	ttctgttaagaatgtgaagaattccataaacaataatttttctggtgaacaaatagaaca	1175
QY	901	ataaagaatatacaataagcagcagcagatgtatacagcaggaacttttataatttt	960
DB	1176	ataaagaatatacaataagcagcagcagatgtatacagcaggaacttttataatttt	1235
QY	961	atacagcaaaagggaataatagatctgtatattatccaacatacaaattagaatactca	1020
DB	1236	atacagcaaaagggaataatagatctgtatattatccaacatacaaattagaatactca	1295
QY	1021	aatccagataaagtagatatagaactatcccatattatttgggttaaatatatac	1080
DB	1286	aatccagataaagtagatatagaactatcccatattatttgggttaaatatatac	1355
QY	1081	atcacaaatcataatgaatttttccaatagaacaatacagaagtgtatagttcccgaaat	1140
DB	1356	atcacaaatcataatgaatttttccaatagaacaatacagaagtgtatagttcccgaaat	1415
QY	1141	aaagaattatagatattggtcaggaaaacaacatacagataaataatcaatatagtg	1200
DB	1416	aaagaattatagatattggtcaggaaaacaacatacagataaataatcaatatagtg	1475
QY	1201	gcaagatgtataggagatagcgatattatagaagaacaattaccgaactgcatcatc	1260
DB	1476	gcaagatgtataggagatagcgatattatagaagaacaattaccgaactgcatcatc	1535
QY	1261	gatgtagaatatttttcttaaaaatatttggaaacatttaattcccggaagtataaacctca	1320
DB	1536	gatgtagaatatttttcttaaaaatatttggaaacatttaattcccggaagtataaacctca	1595
QY	1321	aaagaacctataactgaacgaaccccggaagactgtgagacacataagatgaatattcca	1380
DB	1596	aaagaacctataactgaacgaaccccggaagactgtgagacacataagatgaatattcca	1655
QY	1381	ggactcgttatatccaacataaggagatataatcacaaacaatgatgctatatagaaact	1440
DB	1656	ggactcgttatatccaacataaggagatataatcacaaacaatgatgctatatagaaact	1715
QY	1441	agacttatatcaactctgacccgtgtgtagacagataatgaatgggaactgtgtcaaatga	1500
DB	1716	agacttatatcaactctgacccgtgtgtgtagacagataatgaatgggaactgtgtcaaatga	1775
QY	1501	tgctatgaaaaaataatgcgaatgggaaagcagcccatattataacacgtacataaagaa	1560
DB	1776	tgctatgaaaaaataatgcgaatgggaaagcagcccatattataacacgtacataaagaa	1835

[illegible]

QY 1183 aatcaaaataatagcagatgtataggtatgcatttattatgaagaacaa 1242



```
QY 703 aacttaaatgaatttttcaaatattatataaacttaacttgaagaa 762
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550
QY 765 acatttcggaagaagtaattcaatttcgttcataaagtcgcattcatgc 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 490
QY 823 gtccagagatgcattcttcgttaagaatgaatccataatacaataatttg 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430
QY 883 tctgaacaaataaacaataaagaattacatagggcagcgatgtatcagcaga 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 370
QY 943 accttataataattttaaatacagaagagaaatagatcttgtattatccaaca 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310
QY 1003 tacaatttgaactaacaatccagataatgtagatgtactatccatttat 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250
QY 1063 ttgtgtttaaataacatacacaatcattaggttttcaattagaacagaagt 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 190
QY 1123 ggaattagtcggaataaagaattatagaatttggtagaataaacaatacagataa 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130
QY 1183 aatcaaaataataataggcagatgtagatagatgcatattatgaagaacaa 1242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 70
QY 1243 ttaccaactgcatctcgatgtagatatttcttacaataatttggaaacatttaac 1302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 10
QY 1303 cctgaagta 1311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 WWWWWWWWWWWW 1

RESULT 6
AAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 98US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RW;
```

```
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 12.4%; Score 219; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

QY 523 gaattgaagatcagaagaagtcgcagatgaattatcttcttccctaggtga 582
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 789 GCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 730
QY 583 gaaccatagatagatagaagaagaattgattatgtagataaatgaagctt 642
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 670
QY 643 gtaattgatttgaattagatggtgtagatgtagtgcggaaacacatgggaatttac 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 610
QY 703 aacttaaatgaatttttcaaatattatataaacttaacttgaagaa 762
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 550
QY 765 acatttcggaagaagtaattcaatttcgttcataaagtcgcattcatgc 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 490
QY 823 gtccagagatgcattcttcgttaagaatgaatccataatacaataatttg 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 430
QY 883 tctgaacaaataaacaataaagaattacatagggcagcgatgtatcagcaga 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 370
QY 943 accttataataattttaaatacagaagagaaatagatcttgtattatccaaca 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 310
QY 1003 tacaatttgaactaacaatccagataatgtagatgtactatccatttat 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 250
QY 1063 ttgtgtttaaataatacacaatcattataggttttcttacaataatttggaaacagagt 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 190
QY 1123 ggaattagtcggaataaagaattatagaatttggtagaataaacaatacagataa 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 130
QY 1183 aatcaaaataataataggcagatgtagatagatgcatattatgaagaacaa 1242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 70
QY 1243 ttaccaactgcatctcgatgtagatatttcttacaataatttggaaacatttaac 1302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```

PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match      12.4%; Score 219; DB 22; Length 938;
Best Local Similarity 0.8%; Pred. No. 1.7e-28;
Matches 6; Conservative 569; Mismatches 214; Indels 0; Gaps 0;

QY 523 gaaatagacgttcagaaagtcagtcagatgaatttcttctatccttagtgga 582
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 789 GCMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 730

QY 583 gaaactatgatagatagaaagaaattgattatgtgataaataatgaagctt 642
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 729 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 670

QY 643 gtaatgatttgattagatggtgtagatagtcggaaccacatgggaagtttac 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 669 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 610

QY 703 aacttaagaatlaatttccaattatataataaataaactgttgaagaaa 762
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 609 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 550

QY 763 actatcgaagaagtaattcaattcgtgtcatcaaatggtcatatcatgc 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 549 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 490

QY 823 gttcagagttgacatcttcgtaagaatgccataaacaataattttg 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 430

QY 883 tctgaacaatgaaacaataaagaattcatagggcagcagatgtatcacggga 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 429 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 370

QY 943 actttatcaatttttaatacagaagaagaatagatcttgattatccaaca 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 369 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 310

QY 1003 tacaattagaacaataccaatcagataatagtagatagtactatccattat 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 309 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 250

QY 1063 ttgtgttaaatatacaatcaatcatataggttttccattagaacaatacagagtc 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 249 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 190

QY 1123 ggaattagtcgcgaataaataaagaattatgaattgtagaataaacaatacattgaa 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 189 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 130

QY 1183 aatcaataataatagagcagatgtagagatagtgcatattattgaagaagaaa 1242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 70

QY 1243 ttacaactgagatcattgtagatattttcttacaataatttggaaactttaaat 1302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 10

QY 1303 cctgaagta 1311
   : : : :
DB 9 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 1

```

```

RESULT 9
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN W0200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Unex RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match      12.3%; Score 217.2; DB 22; Length 936;
Best Local Similarity 0.4%; Pred. No. 3.4e-28;
Matches 3; Conservative 569; Mismatches 212; Indels 0; Gaps 0;

QY 4 aatttaaatcatcatatttttaattatagatcatcatctgtatctgcaatccaga 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 62

QY 64 accttgaaagaaataataataatcatcttggaataatacggaaataaataat 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 122

QY 124 aaacatcacaacggaatatacatgcttttccaacttaaatcgaaataatgat 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 182

QY 184 ttgtgaatatgatacttattgcgagatggtgttaacttagaattacaataaat 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 183 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 242

QY 244 aaaaataataataaataatagaaataccacaagaacaattttagagagataaaaa 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 243 WMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 302

QY 304 aggaacaagatattatagcagatactatggttcatcgaaacagtcaagtgatagaca 363

```

[illegible]

CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match	12.3%;	Score 217.2;	DB 22;	Length 936;
Best Local Similarity	0.4%;	Pred. No. 3.4e-28;		
Matches	3;	Conservative 569;	Mismatches 212;	Indels 0;
				Gaps 0;

QY	4	aaattaaataataatcttttaattatagtagccacttgattcttgaaattccaga	63
Db	3	www	62
QY	64	aacttgaaaggaanaaataataataatcatcttgggaataacacgggaataaatal	123
Db	63	www	122
QY	124	aaaactcaacaacggaaatacaatgagctctttcacatcttaactgaataagtaat	183
Db	123	www	182
QY	184	tttgagaatagtactctattcgggagatgggtgtaacctagaatcacaaaaaat	243
Db	183	www	242
QY	244	aaaaataataaataaataatgataaanaacacaagacaaatttagaaggataaaaa	303
Db	243	www	302
QY	304	aggaaacaagtattatagcaggaactagtgtcatcaggaacgtcaaggtatagaca	363
Db	303	www	362
QY	364	aaacatgatgtattcaaccacatggtgcatttatatatattgattgtctgcatt	423
Db	363	www	422
QY	424	aaatgttatatgttgtatcttagacaaatttaatgaaagaaagatcctatlaagaaa	483
Db	423	www	482
QY	484	caaggcttagaatgaacactaagtatgacttaagaattgacgtaccgaaaa	543
Db	483	www	542
QY	544	gtacgtccagatgttaattctttaccttciagtgagagaaacctaatgatagata	603
Db	543	www	602
QY	604	gaaagaagaaltgatctgtagataaataatlgaaactgttaatgatttgatag	663
Db	603	www	662
QY	664	ggctgagatctgacctggaaacacatcgggaagtttaacactaaatgaataattt	723
Db	663	www	722
QY	724	tcaaatattatataatataactcttgaagaaaactatccgagaagaagtta	783
Db	723	www	782
QY	784	attt 787	
Db	783	www 786	

RESULT	11
AAF58257	
ID	AAF58257 standard; DNA; 936 BP.



XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other:

Query Match 12.3%; Score 217.2; DB 22; Length 936;  
Best Local Similarity 0.4%; Pred. No. 3.4e-28;  
Matches 3; Conservative 569; Mismatches 212; Indels 0; Gaps 0;

```
QY 4 aatttaaaatcaatattttaaattatagatcactgtatctcgaattccaga 63
DB 3 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 62
QY 64 accctgaagaaataataataataatcattggaataaaggaataaataat 123
DB 63 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 122
QY 124 aaactcaaaaggaataacatgagctctttcacatcttaactgaataagat 183
DB 123 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 182
QY 184 ttgtagaatatgactctatttcgagatgggtgaactctagaattacaataat 243
DB 183 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 242
QY 244 aaaaataataaaaaatgataaaatcaacaaagaaatttagagagataaaaa 303
DB 243 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 302
QY 304 aggaacaaggtattatagacgataactatggttcacgtgaacgtcaagtgataagca 363
DB 303 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 362
QY 364 aaacatagtatccaacccaatggttcaattatataattgcatctgcgcagt 423
DB 363 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 422
QY 424 aatgttataatgatactagacatttaaggaacaaagatccatttaagaaa 483
DB 423 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 482
QY 484 caagcttagaataatgaacctatggtatgattgaatagacglatagaaaa 543
DB 483 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542
QY 544 gtacgtccagatgaattattcttttaacttaagtgagaacctaatagataata 603
DB 543 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 602
QY 604 gaaaagaataatgatttggataaaaatattgaagctgttaatttgattgat 663
DB 603 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 662
QY 664 ggttagatatgactggagacacatgaggaagttaacacttaataatattt 723
DB 663 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 722
QY 724 tcaattatataatataatgaactgttgaagaaaactatccgagaagaaagta 783
DB 723 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 782
QY 784 attt 787
DB 783 wwwwww 786
```

RESULT 13

AAFS8262 standard; DNA; 936 BP.

XX

AC AAF58262;

XX 24-APR-2001 (first entry)

XX

DE Oligonucleotide D2007.

XX Electron-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Dmek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

XX Example 6; Page 128; 159pp; English.

XX The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

XX Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

QY 4 aatttaaaatcaatattttaaattatagatcactgtatcttgcgaattccaga 63

DB 3 ww 62

QY 64 accctgaagaaataataataataatcattggaataaaggaataaataat 123

DB 63 ww 122

QY 124 aaactcaaaaggaataacatgagctctttcacatcttaactgaataatagat 183

DB 123 ww 182

QY 184 ttgtagaatatgactctatttcgagatgggtgaactctagaattacaaaaataat 243

DB 183 ww 242

QY 244 aaaaataataaaaaatgataaaatcaacaaagaaatttagagagataaaaaa 303

DB 243 ww 302

QY 304 aggaacaaggtattatagacgataactatggttcacgtgaacgtcaagtgataagca 363

DB 303 ww 362

QY 364 aaacatagtatccaacccaatggttcaattatataattgcatctgcgcagt 423

DB 363 ww 422

QY 424 aatgttataatgatactagacatttaaggaacaaagatccatttaagaaa 483

DB 423 ww 482



	Best Local Similarity	0.4%;	Pred No. 3.4e-28;	
	Matches	3;	Conservative	569; Mismatches 212; Indels 0; Gaps 0;
QY	4	aattttaaataatcaattttaaattaaatagatcatcgtatcttgcgaattccaga	63	
Db	3	ww	62	
QY	64	accttgaaaggaaaataataataatcattctgtgggaataatcgggaataataaat	123	
Db	63	ww	122	
QY	124	aaaactcatcaacggaaatatacagtgacttttccaccttaactcgaataatgaat	183	
Db	123	ww	182	
QY	184	tttgtagaatatgagatcttatttcgaggagatgggtgtaactcagattacaaaataat	243	
Db	183	ww	242	
QY	244	aaaaatataataaatatagataaatacacaagaacaatttttagagagatataaaaa	303	
Db	243	ww	302	
QY	304	aggaacaacagtcattatagcagagatactatggttcatacgaacagtcagtagagca	363	
Db	303	ww	362	
QY	364	aaacatatgattatccaacccaatgggtgcattttatatattgatttgcctgcatt	423	
Db	363	ww	422	
QY	424	aatatgtatatatgatactagaccatttaatggaagacaaagattccattaaagaaa	483	
Db	423	ww	482	
QY	484	cccgcttagaatatgaaacctatgctatgcttaatggaattagacgtatcaagaaa	543	
Db	483	ww	542	
QY	544	gtacgcacagatgtaattatcttcatcctaggtggagaacactatagatagata	603	
Db	543	ww	602	
QY	604	gaaaaagaattgattatgtgataaaatatgaaagcttcttaatgattttagattag	663	
Db	603	ww	662	
QY	664	ggtgtagatatgactgggaacacatggtggaagtttacaactaaatgaataaattt	723	
Db	663	ww	722	
QY	724	tcaaatatlatatnaatlaacttgtaagaaaactatccggaagaaagtta	783	
Db	723	ww	782	
QY	784	attt 787		
Db	783	www 786		
RESULT	15			
AAAC89668				
ID	AAAC89668	standard; DNA; 1137	BP.	
AC	AAAC89668;			
AC				
DT	13-MAR-2001	(first entry)		
DE	P. falciparum	chitinase coding sequence SEQ ID NO: 1.		
KM	Malaria;	mosquito; chitinase; fungal disease; parasitic disease;		
KM	veterinary disease;	arthropod pest;		
XX				







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:28:34 ; Search time 94.72 Seconds  
(without alignments)  
4217.770 Million cell updates/sec

Title: US-09-579-383-2

Perfect score: 1764  
Sequence: 1 atgaatttaaatcaat.....tggattaccattacaataa 1764

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/1na/PCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.2	3.5	19124	2 US-08-487-826B-13	Sequence 13, Appl
2	60.6	3.4	7218	1 US-08-232-463-14	Sequence 14, Appl
3	60	3.4	834	4 US-08-998-416-305	Sequence 305, App
4	59.2	3.4	837	4 US-08-998-416-288	Sequence 288, App
5	58	3.3	615	4 US-08-998-416-186	Sequence 186, App
6	58	3.3	665	2 US-08-883-795A-36	Sequence 36, Appl
7	57	3.2	660	1 US-07-991-867B-32	Sequence 32, Appl
8	57	3.2	660	1 US-08-107-755A-32	Sequence 32, Appl
9	57	3.2	660	1 US-08-544-332-32	Sequence 32, Appl
10	57	3.2	1511	1 US-07-991-867B-8	Sequence 8, Appl
11	57	3.2	1511	1 US-08-107-755A-8	Sequence 8, Appl
12	57	3.2	1511	1 US-08-544-332-8	Sequence 8, Appl
13	57	3.2	4810	3 US-08-852-629-11	Sequence 11, Appl
14	57	3.2	4838	3 US-08-852-629-15	Sequence 15, Appl
15	54.4	3.1	5852	1 US-07-867-106-2	Sequence 2, Appl
16	53.6	3.0	3095	6 5231168-1	Patent No. 5231168
17	53.4	3.0	636	4 US-08-998-416-1137	Sequence 1137, Ap
18	53	3.0	19124	2 US-08-487-826B-13	Sequence 13, Appl
19	52.4	3.0	665	2 US-08-883-795A-36	Sequence 36, Appl
20	52	2.9	8920	4 US-08-446-855A-1	Sequence 1, Appl
21	52	2.9	8920	4 US-09-150-741-1	Sequence 1, Appl
22	51.2	2.9	5852	1 US-07-867-106-2	Sequence 2, Appl
23	49.8	2.8	240	1 US-08-628-417-6	Sequence 6, Appl
24	49	2.8	1689	1 US-07-991-867B-41	Sequence 41, Appl
25	49	2.8	1689	2 US-08-544-332-41	Sequence 41, Appl
26	49	2.8	6243	2 US-09-056-075-1	Sequence 1, Appl
27	49	2.8	8457	1 US-07-991-867B-1	Sequence 1, Appl

C 28	49	2.8	8457	2 US-08-544-332-1	Sequence 1, Appl
C 29	48.6	2.8	1066	1 US-08-157-101A-4	Sequence 4, Appl
C 30	48.2	2.7	724	4 US-09-007-119-8	Sequence 8, Appl
C 31	47.8	2.7	828	4 US-08-998-416-538	Sequence 538, App
C 32	47.6	2.7	701	4 US-08-998-416-701	Sequence 701, App
C 33	47.4	2.7	854	4 US-08-998-416-534	Sequence 534, App
C 34	47.4	2.7	2230	3 US-08-844-188-34	Sequence 34, Appl
C 35	47.4	2.7	2447	2 US-09-014-969-14	Sequence 14, Appl
C 36	47.4	2.7	9048	3 US-08-973-273-4	Sequence 4, Appl
C 37	47.2	2.7	731	1 US-08-451-405A-2	Sequence 2, Appl
C 38	47.2	2.7	1683	4 US-09-009-443-6	Sequence 6, Appl
C 39	46.8	2.7	3701	4 US-08-845-258-10	Sequence 10, Appl
C 40	46.8	2.7	3701	4 US-08-990-571-10	Sequence 10, Appl
C 41	46.8	2.7	3701	4 US-08-723-142A-10	Sequence 10, Appl
C 42	46.4	2.6	625	4 US-08-998-416-919	Sequence 19, App
C 43	46.4	2.6	713	4 US-08-998-416-956	Sequence 956, App
C 44	46.4	2.6	1431	4 US-09-316-083-2	Sequence 2, Appl
C 45	46	2.6	1736	3 US-09-182-816-24	Sequence 24, Appl

#### ALIGNMENTS

RESULT 1  
US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhuan  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knodde Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 3.5%; Score 62.2; DB 2; Length 19124;

Best Local Similarity	43.0%	Pred. No. 0.00033	
Matches 304;	Conservative 0;	Mismatches 403;	Indels 0; Gaps 0;
QY 629	aaatatgaagcttgtaattgatcttgcatttgaagtggtgtagatatgcctggaaccac	688	
Db 15360	AAATATAGGTGACCTTGATGCTTTTAAATATATATTTTAAATATATATTTATTTATTA	15419	
QY 689	atcggaagcttacaacttaaaatgaatattttccaatttatctatcaatttaatta	748	
Db 15420	AAAAAGAAATATTAACCAATTTTATTAATAATGAAAAAGAAAAATGAAATATTA	15479	
QY 749	acttgcttaagaaaactatccggaagaagtlaattccaattcttcgttcatcaatg	808	
Db 15480	AAATTTATTTAAATATAAAAAAGAGAAAAATTTTAAAAAATATAT	15539	
QY 809	ctgcattcatcgcttccagaagttgcattcttcgttaagaatgaagatccatata	868	
Db 15540	AAAAATATTAATTAATATTAATTTTGATGATGAAATAAAAATGAAAAAGTTATCA	15599	
QY 869	acactaaatttgcctgacacaatagaacaacaataagaatcacatagggcagcgcg	928	
Db 15600	AAATTTAAAAAATTTTATATAAAAAATGATTAATAAAAAATAAAAACAAGA	15659	
QY 929	tgttcagcggaactttatataatttcaatcacgcgaagaagaataagatcttg	988	
Db 15660	AGAAAAAATACATTTAAAAAATATATATATCATTAATACAAAAAAGAAAAA	15719	
QY 989	tattatccaacacatcacatttagaactcacatccagatataatgtgtagatcgtact	1048	
Db 15720	AAATATATTTTAAAT	15779	
QY 1049	tatccatttatatttggcttcaaatataacatcacatcatataggttttcaattag	1108	
Db 15780	AAAAAAT	15839	
QY 1109	aacataacagaggttgatttagtcccgaaataaagaattatagaattggtgaggaataa	1168	
Db 15840	AAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	15899	
QY 1169	caatcatgatataaaatacaataataatagggcagatggtatagggatagcatttat	1228	
Db 15900	TAAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	15959	
QY 1229	ttatgaagaacaacttaccactgagcatcgcattgcattgatatatttttcaaatattt	1288	
Db 15960	TAAATTTAATTTACATGACACATATACATATATATATATATATATATATATATAT	16019	
QY 1289	ggaacaatttaaatcctggaagttacaactccaaagaacttactata 1335		
Db 16020	CACATTTTACACAT	16066	
RESULT 2			
US-08-232-463-14/c			
Sequence 14, Application US/08232463			
Patent No. 5670367			
GENERAL INFORMATION:			
APPLICANT: DORNER, F.			
APPLICANT: SCHEIFLINGER, F.			
APPLICANT: FALKNER, F. G.			
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS			
NUMBER OF SEQUENCES: 52			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Foley & Lardner			
STREET: 1800 diagonal Road, Suite 500			
CITY: Alexandria			
STATE: VA			
COUNTRY: USA			
ZIP: 22313-0299			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.25
2 CURRENT APPLICATION DATA:
3 APPLICATION NUMBER: US/08/232,463
4 FILING DATE:
5 CLASSIFICATION: 435
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US/07/935,313
8 FILING DATE:
9 APPLICATION NUMBER: EP 91 114 300.6
10 FILING DATE: 26-AUG-1991
11 ATTORNEY/AGENT INFORMATION:
12 NAME: BENT, Stephen A.
13 REGISTRATION NUMBER: 29,768
14 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (703)836-9300
17 TELEFAX: (703)683-4109
18 TELEX: 899149
19 INFORMATION FOR SEQ. ID NO.: 14:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 7218 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25 IMMEDIATE SOURCE:
26 CLONE: PT99pt-F15
27 ;
28 ;
29 US-08-232-463-14

```

```

Query Match: 3.4%; Score 60.6; DB 1; Length 7218;
Best Local Similarity: 3.9%; Pred. No. 0.00058;
Matches 15; Conservative 224; Mismatches 148; Indels 0; Gaps 0.

QY 1363 atagatgaatattccaggactcgttattccaacacatagagatatatacaacaacat 1422
      ||||| ||||| || : : : : : : : : : : : : : : : :
Db 1451 ATGAAAGATTTTGGTCRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392
QY 1423 gatctatagaaaactgatcttattcaatcatcagacttggtgtagacatatgaa 1482
      : : : : : : : : : : : : : : : : : : : : : :
Db 1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332
QY 1483 tgggaacttggtcaagatgctatgataaaatatgcatgaggaaagcccatatt 1542
      : : : : : : : : : : : : : : : : : : : : : :
Db 1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272
QY 1543 aacacgactataaagaagcctctatttatatgaaaaggagacatatattaa 1602
      : : : : : : : : : : : : : : : : : : : : : :
Db 1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212
QY 1603 tgggtgcaacaaggacctcggaagtgcaagcatagatcatcacaaaactagatgca 1662
      : : : : : : : : : : : : : : : : : : : : : :
Db 1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152
QY 1663 tccaattgccaagga tagaagaatgagataaaaatatccaataaacacttagaagta 1722
      : : : : : : : : : : : : : : : : : : : : : :
Db 1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092
QY 1723 gaagaaacatatgacaagaagtgcatt 1749
      : : : : : : : : : : : : : : : : : : : : :
Db 1091 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1065

RESULT 3
; Sequence 305, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp

```

```

1  APPLICANT:  Reblischung, Corline
2  TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
3  TITLE OF INVENTION:  AND USES THEREOF
4  NUMBER OF SEQUENCES:  1152
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  No. 6239264rtis Corporation
7  STREET:  3054 Cornwallis Road
8  CITY:  Research Triangle Park
9  STATE:  No. 6239264th Carolina
10 COUNTRY:  USA
11 ZIP:  27709
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  Patentin Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/998,416
20 FILING DATE:  24-DEC-1997
21 CLASSIFICATION:  435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER:  CH 0016/97
24 FILING DATE:  31-DEC-1996
25 ATTORNEY/AGENT INFORMATION:
26 NAME:  Meigs, J. Timothy
27 REGISTRATION NUMBER:  38,241
28 REFERENCE/DOCKET NUMBER:  PE/5-30306/A/CGC1976
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE:  919-541-8587
31 TELEFAX:  919-541-8689
32 INFORMATION FOR SEQ ID NO:  305:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH:  834 base pairs
35 TYPE:  nucleic acid
36 STRANDEDNESS:  single
37 TOPOLOGY:  linear
38 MOLECULE TYPE:  DNA (genomic)
39 ORIGINAL SOURCE:
40 ORGANISM:  PAG1250RP
41
42 US-08-998-416-305

```

	Query Match	3.4%;	Score 60;	DB 4;	Length 834;	
	Best Local Similarity	45.9%;	Pred. No. 0.0005;			
	Matches	275;	Conservative	0;	Mismatches 321;	Indels 3; Gaps 2
QY	478	agaacacagcgcttagaatgaacaacctatggtaagtcgctaataagaattagacgtalc	537			
Db	114	AAATTAATTAGAAAACCATAGTTAATGATTCATTAAGAAMAAATGGAAATTTTGTGC	173			
QY	538	agaaaagcagclccagatgtaatttctttatcccttagctgggaaacc--atatga	595			
Db	174	ATCTTAATTTTTATTTATTTAAATGATTTATTCATCATTAAACAATAAACATTTTAAATGT	233			
QY	596	tagatatgaagaagaataatgatlatcgtgataaataatgaagcttgtaatgattcgc	655			
Db	234	TATATAAATAAATFAGAAATTTACTTATAGAAATTTTATTAATAAGATTTAATTTAATTTT	293			
QY	656	atttagatgctgtagatcatctgactgaggaaaccaatgysgaagllttaacaactaaatgaat	715			
Db	294	AATTTAAATATACCAATTTTATTTATTTAAATGAATTAATTAAGTTTATTTAATTTAAGTGA	353			
QY	716	taaattttcaatttatatatlaaattaatnaactctgttaaagaaaacatctccgaag	775			
Db	354	TATATAATTTAATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA	413			
QY	776	aaaagttaattcaattcttgltcatcaanaabgcgcatalcatgctgttcaagagttg	835			
Db	414	TGTAACCTTGCAATAATTTATTTTATTT-AGHCGTAGTAAATTTTCTATTATTAAGTCAC	472			
QY	836	cattctttctgtaagaatgaaagatctcatataaacaactaaattlctgtcgaacaatbg	895			
Db	473	CCTTAATTTGGAATATTACTCCACTTAATAATTTATTCGPAATATATTTATTTAAGAAATTC	532			

QY 896 aacacaataagaatttcacatagagcagcagcgatgttacacagaagaaactttattata 955  
 Db 553 TTAAATCTAAATAATTTATATCTAAAGTAAATATAATTAATTAATTCCTTTTATTATTATA 592  
 OY 956 tttaataacagcagaagagaaatagatctgtattattcaacatacaatttagaa 1015  
 Db 553 TTATATATATTTATATAGTAATAATTAATTTATTTTATTTTATACATAATTTTTTGATTA 652  
 OY 1016 ctacaacacagatataatgttagatatactactacccattataattggtttaaa 1074  
 Db 653 TAAATATATCCATTAAATGTGTAATTTATTAATAAATATCCTTTAATGATTTAATGATA 711

```

1 RESULT 4
2 US-08-998-416-288/c
3 ; Sequence 288, Application US/08998416
4 Patent No. 6239264
5 GENERAL INFORMATION:
6 APPLICANT: Philippsen, Peter
7 APPLICANT: Pohlmann, Rainer
8 APPLICANT: Steiner, Sabine
9 APPLICANT: Mohr, Christine
10 APPLICANT: Wendland, Jürgen
11 APPLICANT: Knechtle, Philipp
12 APPLICANT: Rebschunig, Corline
13 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII
14 TITLE OF INVENTION: AND USES THEREOF
15 NUMBER OF SEQUENCES: 1152
16 CORRESPONDENCE ADDRESS:
17 ADDRESSEE: No. 6239264artis Corporation
18 STREET: 3054 Cornwallis Road
19 CITY: Research Triangle Park
20 STATE: No. 6239264ch Carolina
21 COUNTRY: USA
22 ZIP: 27709
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Floppy disk
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: Patentin Release #1.0, Version #1.30
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/998,416
30 FILING DATE: 24-DEC-1997
31 CLASSIFICATION: 435
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: CH 0016/97
34 FILING DATE: 31-DEC-1996
35 ATTORNEY/AGENT INFORMATION:
36 NAME: Meigs, J. Timothy
37 REGISTRATION NUMBER: 38,241
38 REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: 919-541-8587
41 TELEFAX: 919-541-8689
42 INFORMATION FOR SEQ ID NO: 288:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 837 base pairs
45 TYPE: nucleic acid
46 STRANDEDNESS: single
47 TOPOLOGY: linear
48 MOLECULE TYPE: DNA (genomic)
49 ORIGINAL SOURCE:
50 ORGANISM: PAG1241RP
51 US-08-998-416-288

```

Query Match	3.4%	Score 59.2	DB 4	Length 837
Best Local Similarity	45.1%	Pred. NO. 0.00073		
Matches 259	Conservative 0	Mismatches 313	Indels 2	Gaps 1
Oy	492	agaataaactatggtatcga	tcctcattgaattgaacgtatcagaagaagtacgcc	551



```

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
City: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 665 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 32
US-08-883-795A-36

Query Match 3.3%; Score 58; DB 2; Length 665;
Best Local Similarity 45.8%; Pred.No. 0.0012;
Matches 241; Conservative 0; Mismatches 280; Indels 5; Gaps 1.

Db 585 aacctatgatgatagatagagaaaaagaatcgatctatgtggatataaatatggaagctgt 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 AAAATATTGAATTTATAAATATGTAATTTAAATCTTAATTATATAAAATATGTAATAT 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 645 taatgatttgattagatggtgtagatattgacgtcggaaccacacgtgggaagtttaca 704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 AAATACCTTAATTTATAAATATGTAATTTATTAATTAATCTTTATATAAAATATGTAAT 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 705 cttaaatgaattaaatttccaattatataattaataaattaaactgtttaagaanaaac 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 TATGTAATTTATAAACATTTTAAATTAATAAAATATGTAATTTATTAACATTTTATTA 382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 765 tatccggaagaaaagtttaattccaatttcggttcacatcaaatgctgcatatcatgcgt 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 TATGTAATTTATAAACATTTTAAATTAATAAAATATGTAATTTATTAACATTTTATTA 322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 825 ttccggagctgcaccttctcgtlaaagatgaaagatcccatataaacactaaatttgc 884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 TATGTAATTTATAAACATTTTAAATTAATAAAATATTTAATTAATAACATTTTATTA 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 885 tgaacaataagaaaacaataaagaattacacagggcagcagcagatgltacacagcagaac 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TATTTAATTTATAAATATTTTAAATTAATAAAATATTTAATTAATAAAATATTTAATTA 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 945 t-----ttatctaataatttctaatacagcagaaggaataatagactctgtattatcca 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 TATTTAATTTATAAATATTTTAAATTAATAAAATATTTAATTTAATAAAATATTTAATTA 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1000 acatacaatttgaagaacacacacacccagatataatgtagatagtactatcccatcta 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 TATTTAATTTATAAATATTTTAAATTAATAAAATATTTAATTAATAATTTTAAATTAATA 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Oy 1060 tatttggtttaaataatacaatccatcaactaggttttcac 1105
      ||||| + ||| + ||| + ||| + ||| + ||| + ||| + |||
Db 81 TATTTAATTATRAAATTTTAAATTAATTAATTAATTAATTAAT 36

RESULT 7
US-07-991-867B-32/c
: Sequence 32, Application US/07991867B
: Patent No. 5476781
: GENERAL INFORMATION:
:   APPLICANT: Moyer, Richard W.
:   APPLICANT: Hall, Richard L.
:   APPLICANT: Gruidl, Michael E.
:   TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
:   NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: David R. Saliwanchik
:   STREET: 2421 N.W. 41st Street, Suite A-1
:   CITY: Gainesville
:   STATE: FL
:   COUNTRY: USA
:   ZIP: 32606
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/07/991,867B
:   FILING DATE: 12-DEC-1992
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: WO 92/14818
:     FILING DATE: 12-FEB-1992
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 07/827,685
:       FILING DATE: 30-JAN-1992
:       PRIOR APPLICATION DATA:
:         APPLICATION NUMBER: US 07/657,584
:         FILING DATE: 19-FEB-1991
:         ATTORNEY/AGENT INFORMATION:
:           NAME: Saliwanchik, David R.
:           REGISTRATION NUMBER: 31,794
:           REFERENCE/DOCKET NUMBER: Uf114.C3
:           TELECOMMUNICATION INFORMATION:
:             TELEPHONE: 904-375-8100
:             TELEFAX: 904-372-5800
:           INFORMATION FOR SEQ ID NO: 32:
:             SEQUENCE CHARACTERISTICS:
:               LENGTH: 660 base pairs
:               TYPE: nucleic acid
:               STRANDEDNESS: double
:               TOPOLOGY: unknown
:             MOLECULE TYPE: DNA (genomic)
:             US-07-991-867B-32

Query Match 3.2%: Score 57; DB 1; Length 660;
Best Local Similarity 45.9%; Pred. NO. 0.0019;
Matches 265; Conservative 0; Mismatches 310; Indels 2; Gaps 2;

Oy 706 ttaaatgaaataatttccaataatatataatcaatcaactgtgtaagaanaaac 765
      ||| + ||| + ||| + ||| + ||| + ||| + ||| + ||| +
Db 629 TTTAAATATATATATATTTCTTTATATGATGATTAATGACATTAATGATTAATCTTTT 570

Oy 766 atcccgagaagaagttaattcaatttcgtgtcacaatgcgtcatcatgcgtc 825
      || + || + || + || + || + || + || + || + || + ||
Db 569 GGAATTTTATCAAAATGATATATATTTTATTTTAAACATGAAAGATATATCTAAATTTT 510

Oy 826 tcgaagagtcgatcttcttcgtgaagaatgagaagaatcctacataaacacaaatttgcct 885
      || + || + || + || + || + || + || + || + || + ||
Db 509 AAAATTAATTAATTTTCTATTAATTAATTTATTAATCTTTTATTTGATTTGATTCATATTTAAC 450

```



OY	886	gaacaaatagagaaccaataaagaattcataggcgagcga tgcattcacgcaagaact	945
Db	449	TCAACTATACGTAATAAGAAAAACCTTTATTAAAGTATACCAATTTTTATTAGATATTAT	390
OY	946	tctaatcatcttttaatatacgcaagagaaaaataga- ccttgta tttatccaacata	1006
Db	389	AATTTTTTTAAATTACTAAATTATTAAATAAAATTATAGCATTTTATTATTCATGTGATTA	330
OY	1005	caatttagaacctaacaatccogatalaetlgtlga atgat tactta tccat tca tt	1066
Db	329	CAATTTAAAAATTCTATATAGAATGGTAGATATAATATCTTATTTGCATTTTTTGTTA	270
OY	1065	tggtttaaatalataacatccaalcalatltag ltt tca tta tga aca ta e ag t gg	1122
Db	269	TAAAGATATATCTAAATATGTTATTTTAAATTTGTTATAAAATTTAAATTAATATA	210
OY	1125	attatg tc ccg aaa ta a ga aat t at aga att t gg ta g ga aaa aca ta c t ga ta aaa	1186
Db	209	TTTAAATTTGAATAATTTAAACITTTTAATTTTTCGAAATTTATTTTAAATTTATATTA	150
OY	1185	tcaaataataatataatggcagat tgt ata a agg ata at gg catt t t ta t ba aa ga aac aat	1244
Db	149	TCATAAAT-AATTTTATATGCAATTTCTTAATTAATTAACAAATTTTTTAAATATATTAATATT	91
OY	1245	accac tgc gat cat tgc gat tgc at t t t t c ta	1281
Db	90	AATTAACAATTATCTCTGTTTATTATTATTTTTTTTAA	54

RESULT 8  
US-08-107-755A-32/C  
Sequence 32, Application US/08107755A  
Patent No. 5721352  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF114.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 660 base pairs  
TYPE: nucleic acid

```

; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-107-755A-32

```

Query Match	3.2%;	Score 57;	DB 1;	Length 660;
Best Local Similarity	45.9%;	Pred. No. 0.0019;		
Matches 265;	Conservative	0;	Mismatches 310;	Indels 2;
			Gaps	21

QY	706	tttaatgaataaatttttccaattatattatcaatttaatttaactgttgaagaaact	765
Db	629	TTTAAATATATATATATTTCTTATATAGATGTAATTCATTAATAGACTTTACTTTT	570
QY	766	attccggaagaaagttaattccaattctcgtgtccaaatgctcatcaccgct	825
Db	569	GGAAATTTTATCAATAGTATATATTTTATTAACATGGAAGATATATCTAATTTT	510
QY	826	tcgaggttcgcatcttcctcgttaagaatgaaatccatcatabaacactaatttgcct	885
Db	509	AAATTAATTAATTTTCTATTAATTTATTAATCTTTATTTGATTCGATTCATTTTAC	450
QY	886	gaacaataagaacaataaagaattacataggcgacgagatgctacagcagaact	945
Db	449	TCAACTATACTAATAGGAAAAACATTTTAAAGTTCCCAATTTATTTTATATTTT	390
QY	946	tttaataatltttaacagcaagaagagaaataga-ctctgtatattatccaacata	1004
Db	389	AATTTTAAATTTTACTTAATATTAATTAATTAATTAATTAATTAATTAATTAATTA	330
QY	1005	caatttgaagaactcaaatccagatataatagtgagatagtactatcccatlatalt	1064
Db	329	CAATTTAAAAATTCATAAGATGTGGTAGATTAATTAATTAATTAATTAATTAATTA	270
QY	1065	tgttttaaatataaacctcaacatcatatagggttttcaattagaacaatacagaagcgg	1124
Db	269	TAAATTTATTCATAATTTGTTATATTTTAAATTTTGTATTAATTAATTAATTAATTA	210
QY	1125	atttagtcccgaaaaataaagaattatagaaattgtgtaggagaaacaatacatcagataaaa	1184
Db	209	TTTAAATTTGAAATATATTAACCTTTTAATTAATTTTTCGAAATTTAATTAATTAATTA	150
QY	1185	tcaaaataataataggcgagatgctatagggatatgcatcttattatgaaagaacaatt	1244
Db	149	TCAATAAAT-ATAATTAATATGCAATTTCTCTATAATTAACATAATTTTAAATTAATTAAT	91
QY	1245	accacatgagatcgcgatgtagaatttttcttaca	1281
Db	90	AATTAACATTAATCTGTTTATTAATTAATTTTAAAA	54

RESULT 9  
 US-08-544-332-C  
 Sequence 32, Application US/08544332  
 Patent No. 5935777  
 GENERAL INFORMATION:  
 APPLICANT: Moyer, Richard W.  
 APPLICANT: Hall, Richard L.  
 APPLICANT: Gruidl, Michael E.  
 TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System  
 NUMBER OF SEQUENCES: 77  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Gerard H. Bencen  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25



	Matches	265:	Conservative	0:	Mismatches	310:	Indels	2:	Gaps
Oy	706	ttaaatgaattcaatctttccaatatattataattaacttgaagaaact	765						
Db	1480	TTTTTATTAATTATTTCTTTATAAGATGATAATCCAAATGAAGCTATACCTTTT	1421						
Oy	766	attccggaagaagttaattccaatttcctgggtcacaacgcgcattacalcgcgt	825						
Db	1420	GGAATTTTATGCCAANTGTATATTATTTTTTTTAACTTGAAAGCATATCTAATTTTTT	1361						
Oy	826	tcaaggagtgcactcttcctgtaaagatgsaagactccataaacactaaatlctc	885						
Db	1360	AAATTAATTAATTTTTCATTAATTAATTAATTAATTTTATTTGATGATGCATTAATTTAAC	1301						
Oy	886	gaaccaataagaacaataagaattaccatagggcgagcgatgltacacaggaact	945						
Db	1300	TCAACTATATCTAATATGGAAAACATTATTTAAAGTTACCAAATTTATTTTAGATATTATT	1241						
Oy	946	ttatataattctttaatcacagcaagaagaaaataga-tcttglatattccaacata	1004						
Db	1240	AATTTTTTAAATTTTACTTAATTTATTAATTAATTAATTAATGTCATTTATTTAATGATGTCA	1181						
Oy	1005	caatttagaactacaatccagatataatggtlagaatalgtacttaoccatttat	1064						
Db	1180	CAATTTAAAAAATTCATACAATAGTGTACTTAATAATTTCTAATTTGCTAATTTTGGTTA	1121						
Oy	1065	tgtttaaataatacatcacatcatattaggttttccattagaacaacaagsgtgg	1124						
Db	1120	TAAAGTATATCTAAATATGCTATTTATTTTAAATTTTGTTATATAAATTTAAATTAATAA	1061						
Oy	1125	attatgcccgaaaaataaagaattatcgaattggltaggaaaaacatacalgataaaa	1184						
Db	1060	TTTTAAATTTGAAATATATTAACCTTTTAATATTTTCTCGAATATTTATTTAAAAATATTATTA	1001						
Oy	1185	tcaaaataataatagggcgatggtatagggatagcgcatattatitagaagaacaatt	1244						
Db	1000	TCATATAT-ATATATATATGCAATCTTCCTAATTAATTAACTAATTTTTTTAAATATATTAAATTT	942						
Oy	1245	accaactgcatcatcgatgbtagatatttcttaca	1281						
Db	941	AATAACATTATCTCTGTTTATTTATTTATTTTAAAA	905						
<b>RESULT 11</b>									
US-08-107-755A-8/c									
Sequence 8, Application US/08107755A									
Patent No. 5721352									
GENERAL INFORMATION:									
APPLICANT: Moyer, Richard W.									
APPLICANT: Hall, Richard L.									
APPLICANT: Gruidl, Michael E.									
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System									
NUMBER OF SEQUENCES: 40									
CORRESPONDENCE ADDRESSES:									
ADDRESSEE: David R. Saliwanchik									
STREET: 2421 N.W. 41st Street, Suite A-1									
CITY: Gainesville									
STATE: Florida									
COUNTRY: U.S.A.									
ZIP: 32606									
COMPUTER READABLE FORM:									
MEDIUM TYPE: Floppy disk									
COMPUTER: IBM PC compatible									
OPERATING SYSTEM: PC-DOS/MS-DOS									
SOFTWARE: Patent Release #1.0, Version #1.25									
CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US/08/107,755A									
FILING DATE: 19-AUG-1993									
CLASSIFICATION: 435									
PRIOR APPLICATION DATA:									
APPLICATION NUMBER: US 07/827,658									
FILING DATE: 30-JAN-1992									
PRIOR APPLICATION DATA:									

```

1  APPLICATION NUMBER:  US 07/657,584
2  FILING DATE:  19-FEB-1991
3  ATTORNEY/AGENT INFORMATION:
4  NAME:  Saliwanchik, David R.
5  REGISTRATION NUMBER:  31,794
6  REFERENCE/DOCKET NUMBER:  UFI114,C2
7  TELECOMMUNICATION INFORMATION:
8  TELEPHONE:  (904) 375-8100
9  TELEFAX:  (904) 372-5800
10 INFORMATION FOR SEQ. ID NO.:  8:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH:  1511 base pairs
13 TYPE:  nucleic acid
14 STRANDEDNESS:  double
15 TOPOLOGY:  unknown
16 MOLECULE TYPE:  DNA (genomic)
17 ORIGINAL SOURCE:
18 ORGANISM:  Amsacta moorei entemopoxvirus
19 FEATURE:
20 NAME/KEY:  CDS
21 LOCATION:  complement (18..218)
22 FEATURE:
23 NAME/KEY:  CDS
24 LOCATION:  complement (234..782)
25 FEATURE:
26 NAME/KEY:  CDS
27 LOCATION:  852..1511
28
29 US-08-107-755A-8

```

[illegible]



```

? REFERENCE/DOCKET NUMBER: UF-184
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 407-426-7500
? TELEFAX: 407-839-8589
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4810 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cDNA
? HYDROTHERMAL: NO
? ANTI-SENSE: NO
? US-08-852-629-11

```

Query Match	3.2%;	Score 57;	DB 3;	Length 4810;
Best Local Similarity	45.9%;	Pred. No. 0.0029;		
Matches 265;	Conservative	0;	Mismatches 310;	Indels 2; Gaps 2.

[illegible]

RESULT 14  
 US-08-852-629-15/C  
 ; Sequence 15, Application US/08852629  
 ; Patent No. 6106825  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Moyer, Richard W  
 ; APPLICANT: Li, Yi  
 ; APPLICANT: Hall, Richard L  
 ;  
 ; TITLE OF INVENTION: ECTOMORPOXVIRUS-VERTEBRATE GENE DELIVERY  
 ;  
 ; TITLE OF INVENTION: VECTOR AND METHOD  
 ;  
 ; NUMBER OF SEQUENCES: 17  
 ;  
 ; CORRESPONDENCE ADDRESS:

ADDRESS: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,629  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Bencken, Gerard H  
REGISTRATION NUMBER: 35,746  
REFERENCE/DOCKET NUMBER: UF-184  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 407-426-7500  
TELEFAX: 407-839-8589  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4838 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

Query Match	3.28;	Score 57;	DB 3;	Length 4838;
Best Local Similarity	45.98;	Pred. No. 0.0029;		
Matches 265; Conservative	0;	Mismatches 310;	Indels 2;	Gaps 2

[illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2002, 14:24:29 ; Search time 1543.83 Seconds

(without alignments)  
12278.265 Million cell updates/sec

Title: US-09-579-383-2

Sequence: 1 atgaatttaataatcaat.....tgaatttaacattacaataa 1764

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	104.4	5.9	1101	13	AL061936 Drosophila
2	89.8	5.1	1101	13	CNS00021J
3	89.2	5.1	1101	13	CNS000EVL
4	84.6	4.8	1101	13	CNS00039G
5	83	4.7	1101	13	CNS000EVL
6	82.8	4.7	1101	13	CNS000EVL
7	81.2	4.6	945	13	CNS04DOK
8	79.8	4.5	994	13	CNS04NOJ
9	79.4	4.5	1101	13	CNS000KB5
10	79.4	4.5	1225	13	CNS0161D
11	79.2	4.5	1161	13	CNS073Y8
12	78.4	4.4	1101	13	CNS0006J

C 13	77.6	4.4	524	13	CNS01090	AL167541 Tetraodon
C 14	77.4	4.4	876	13	CNS00961	AL053529 Drosophila
C 15	77.2	4.4	839	11	BF259858	BF259858 HVSMEF002
C 16	77.2	4.4	1190	13	CNS020N7	AL206908 Tetraodon
C 17	77.2	4.4	879	13	CNS01JRG	AL147405 Anopheles
C 18	77	4.4	963	13	CNS00XAL	AL054918 Drosophila
C 19	76.8	4.4	1007	13	CNS06X95	AL419462 T3 end of
C 20	76.8	4.4	1101	13	CNS00E07	AL069440 Drosophila
C 21	76.6	4.3	942	13	CNS018GS	AL109318 Drosophila
C 22	76.4	4.3	1201	13	CNS0167M	AL106396 Drosophila
C 23	76.2	4.3	1101	13	CNS003B4	AL064082 Drosophila
C 24	76	4.3	1337	5	BF630719	BF630719 HVSMEB001
C 25	75.6	4.3	928	13	CNS00DXY	AL071865 Drosophila
C 26	75.2	4.3	1091	13	CNS014AC	AL103902 Drosophila
C 27	75	4.3	816	13	CNS02KJY	AL120164 Tetraodon
C 28	74.8	4.2	1009	13	CNS00587	AL057153 Drosophila
C 29	74.6	4.2	878	10	AL536095	AL536095 AL536095
C 30	74.6	4.2	971	13	CNS06D25	AL1933267 T3 end of
C 31	74.6	4.2	1101	13	CNS003BD	AL064091 Drosophila
C 32	74.6	4.2	1125	10	AL547503	AL547503 AL547503
C 33	73.4	4.2	1101	13	CNS0039G	AL063921 Drosophila
C 34	73.4	4.2	1101	13	CNS003B0	AL064078 Drosophila
C 35	73.4	4.2	1101	13	CNS00LJT	AL068307 Drosophila
C 36	73.2	4.1	829	13	CNS03LUB	AL249932 Tetraodon
C 37	73.2	4.1	1146	13	CNS021G2	AL176843 Tetraodon
C 38	73	4.1	1101	13	CNS01V2	AL108516 Drosophila
C 39	73	4.1	1151	11	BC309087	BC309087 HVSMEC000
C 40	72.8	4.1	839	13	CNS01LXK	AL100395 Drosophila
C 41	72.8	4.1	1101	13	CNS0182P	AL108811 Drosophila
C 42	72.8	4.1	1101	13	CNS058GV	AL1325912 Tetraodon
C 43	72.6	4.1	1101	13	CNS00KAE	AL077628 Drosophila
C 44	72.4	4.1	961	13	AL691838	AL691838 ENTW156TF
C 45	72.2	4.1	922	13	AZ548363	AZ548363 ENT207TF

#### ALIGNMENTS

RESULT 1  
CNS0021J 1101 bp DNA GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL061936 GI:4940214  
VERSION AL061936  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
TITLES  
AUTHORS  
JOURNAL  
COMMENT

Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila genome project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

[illegible][illegible]

	Matches	136;	Conservative	302;	Mismatches	263;	Indels	5;	Gaps	2;
Oy	18	aataattttaatatgaatcatccttgtagtatttcgcaaatccacaacgttgaaagaaa	77							
Db	1097	RWGGDTTWDRDTRKKDDWDMTKMWTKDKRADDRRMGAGDDRMAWDGAGTWTATYATMWMWW	1038							
Oy	78	aaataabataataaacatcgtaggatataacggagaataaaaaataaacatcacaaac	137							
Db	1037	WWATTMDTYMDKMMWWMATAKTDYTATMYRTAMRADWAGDRACGRDRNRATADAGAGR	978							
Oy	138	ggaatatagtaagtcctttlccaactccta--cgaaataagttaatlttgttagaat	195							
Db	977	DGGRKRKKRKRRKDDBDKRGKGKKKAARAAKWARKWMDDMWMDMWDGAKDKRRADD	918							
Oy	196	ggatcttatlcgcggagatagggtgaacctgaagattccaaaaataatcaanaataat	255							
Db	917	DGAGCKDDDGSKRDADDOTDGTGKDDDDKDKMDWKAKAGTWGDATWAANAATITMWMGNAD	858							
Oy	256	aaaaatgatalagaaaaccaccaagaacaatlittagegagatlaaaaaaagaaacagyt	315							
Db	857	ADMWTVDAADAADDWMDRWMAWKMVNDAMAAGARTADRBDWGDAAGKRGARRRRDRR	798							
Oy	316	attatagaagatalectatgyttcatacgagaacgccaaglygataagsccaacatatglt	375							
Db	797	AADKRRAADDRDAATYWTTTTRTDTPDDMKWKTDTYWRMAADRTWDRDDDDDKDRAGT	738							
Oy	376	gattcaaacccaatgtytcaattatataatcatgcatltycccgcattaataatgtlatat	435							
Db	737	AGRKMRFTWRKRRKRRORRWMDADADPTARODRRRRGGDAGAGKKTGKRRRRDRAT	678							
Oy	436	gatgtatctagacccattaaatygaaagacaagaattccattatgaagaacaagcgttagaa	495							
Db	677	WDRDIDAWAADAAWMTTPTDTPDDMKRRDRKRGARRRRRTTRAAMDWMTWAMWAKMD	618							
Oy	496	tatgaaccclatgytatgcatgctaagaatgaagcatgaagatcaaagaagtagctccogat	555							
Db	617	WKT RAD RMDRNAADDTWTARAKRADRMWARARARARARDARARABRRWTYKCTTATYAT	558							
Oy	556	glaatattcctttatcctlagtggtaggaagaacctaatgatagatatagataaaaaaagaa	615							
Db	557	TWAAAANNAWAWAATAATTAATTTTTTTW---TTTTTTTTTTTWTAAWMAAAWMTATW	501							
Oy	616	gattatgtagttaaataatltgaagctlltgaagtlitgaagtlitagatgaglytagaat	675							
Db	500	AAWTAANAAAAAAMAAWAAATTTTTTTTTTTTWTAAWMTATWTTTWTITTMNAATT	441							
Oy	676	gactlggaaccacatcgggaagtlttcacactaaatgaatgaat	721							
Db	440	TTTTWTTTWNAAATTTATTTTTTTTAAAAAAMMMMMWTTATATX	395							
RESULT	4	CNS00EVL	1101 bp	DNA	GSS	04-JUN-1999				
Locus	DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC:								
		BACR2B23 of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.								
Accession	VERSION	AL069706								
SOURCE	KEYWORDS	AL069706.1 GI:4949849								
ORGANISM		GSS.								
		fruit fly.								
		Drosophila melanogaster								
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;								
		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;								
		Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
REFERENCE		1 (bases 1 to 1101)								
AUTHORS		Genoscope.								
FILE		Direct Submission								
JOURNAL		Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :								
		BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr								
		- Web : www.genoscope.cns.fr)								
COMMENT		Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).								

## FEATURES

BASE COUNT	419 a	91 c	60 g	299 t	232 others
ORIGIN					

Matches 205; Conservative 118; Mismatches 297; Indels 3; Gaps 1;

CC-0

382 aaaccaatggtgtcaatttataattgcatttgcctgcattaatatgtatgta 441

[illegible]

DD 616 AATTGTTTAAATWAWATTATTAATAAATTTWTAAATTTATTTWW 557

fly), genomic survey sequence.  
AF078714

REFERENCE 1 (bases 1 to 1101)

COMMENT  
Determination of this BAC-end sequence was carried out as part of a collaboration with the National Research Council (NRC) and the University of Alberta.

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## ORIGIN

Matches 222; Conservative 102; Mismatches 306; Indels 1; Gaps 1,

[illegible]

QY 310 caaggtattatagcaggaatactatggttcattggaacagtcgaagtgatagagcaaacat 365

Db 525 TTTAATTWTTWATTAATATTTTTFKKWTTWKATKTDWTKKAKTAAKTAAWAAWTTT 584

Db 863 WATAKAKWGTWCAKWTADKKDTGGTGTGGWGTGWRGRGTGKKRRGRACGDTRGDRTRT 804

```
/clone="101H21
/clone_11b="G"
```



```

RESULT      9
CNS00K85   1101 bp   DNA   GSS   03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BAC39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL077453
VERSION    AL077453.1 GI:4956930
KEYWORDS   GSS.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : segreif@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
    source
        1..1101
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RPCI-98"
        /clone="BAC39P05"
        /note="end : 17"

BASE COUNT  540 a      0 c      139 g      252 t      170 others

ORIGIN
Query Match      4.5%; Score 79.4; DB 13; Length 1101;
Best Local Similarity 41.2%; Pred. No. 0.01;
Matches 282; Conservative 28; Mismatches 370; Indels 5; Gaps 1;

Oy  54  aaattccagaaccttgaaagaaataataataataatcttggaataacaggga 113
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   5  AAAAAAAAAAAAAAAAAAATTAANANNNANNNAAAAAAAAAAAAAAAAAAAA 64

Oy  114 aaataaaataaactcaacacggaataacatgagctcttccatccttaaatcgaa 173
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   65 AAAAAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 124

Oy  174 taatagtaatttgtaataatgatactatgacgagatggtgttaactctagaattac 233
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   125 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAATT 184

Oy  234 aaaaataataaataataataaataatgatagaataacacgaagaattttagagga 293
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   185 TAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 244

Oy  294 gataaaaaaaggaacaggtatatacgacgatactatggttcagtgaacagtcagg 353
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   245 AAAAAAAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 304

Oy  354 tgaatagcaaaaatgattgattcaaccacaaatggtgcacatttatattatgcat 413
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

Db   305 ATAAAAAAAAAAAAAAAAAAAAAAAAANNNNNNAATTAATAAATAAATAATTATA 364

Oy  414 tgcctgcattaataatgatatatglatctagacacatttaaggaagacaagatcct 473
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   365 ATATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 419

Oy  474 attaagaaaacacgccttggaataataaacctatgltatgactcttaataattgaac 533
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   420 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 479

Oy  534 taccagaagaagtcagtcacagatgatacttcttacccttagtcggaagaaccata 593
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   480 AAAAAAAAAATTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 539

Oy  594 gatagatatagaagaagaattgatatgtgataaataatgaagcttgtaattgat 653
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   540 AAAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 599

Oy  654 tgattagatggtgtgatatgactgacggaacacacatggaagtttacaactaatga 713
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   600 TAAATTTTTRTKRKGGAATTTKGGGTTGTTTWTATKCTGGTGGKWTTRKGG 659

Oy  714 attaaatttcaaatattatatt 738
    111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db   660 KADTTAGTKTGCTTTTATATATK 684

RESULT      10
CNS0161D   1225 bp   DNA   GSS   26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL106171
VERSION    AL106171.1 GI:5620504
KEYWORDS   GSS.
SOURCE     fruit fly.
ORGANISM   Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1225)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL     Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : segreif@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelOBAC11.

FEATURES
    source
        1..1225
        /organism="Drosophila melanogaster"
        /plasmid="pBelOBAC11"
        /db_xref="taxon:7227"
        /clone_lib="DrosBAC"
        /clone="BACN15C18"
        /note="end : SP6"

BASE COUNT  266 a      128 c      38 g      368 t      425 others

ORIGIN
Query Match      4.5%; Score 79.4; DB 13; Length 1225;
Best Local Similarity 29.8%; Pred. No. 0.0099;
Matches 177; Conservative 139; Mismatches 277; Indels 1; Gaps 1;

Oy  646 aatgatttgatttagatggtgtgatatgactggaacacatggaagtttacaac 705
    :: 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```



Db	1224	MWTTATGATATGTCTTTCATTTTTTTWWAMAMAMMMMAAAMAAAMWTCTTTTAATAAAAT	1165
OY	706	ttaaatgaaatlaattttcaaa-ttatatatlaaatlaataactgttaagaanaac	764
Db	1164	ATATATATATATATTATATWWWAAAAMMTTTTTTTATATWATTTATMATAAAAAAMAAATATTAAMW	1105
OY	765	tattccgaagaaaagtlaattccaattcttgcttccaatgcgatcatatcgcgt	824
Db	1104	MAMMMMAAMAMAMAMAMAMATMTWTWTWKATATATATAARTRMATATATAAAATATAMAMMW	1045
OY	825	ttcaggagltgcattcttcgttaagaagalcotcalatacacctaattlctgc	884
Db	1044	TVMMAAAATATATTTTTTTTTTTTTTMAAAAAAAAAAAAAAMAPATAATATMTATMAAMMW	985
OY	885	tgcacaataagaacaataagaatatacaggcgacagcgatggtatcacgaagAAC	944
Db	984	TAAAAAAAATATATWATATTTTMAAMWTTTTTTTAAAMWTTTTTTTTTTTWTTTAAAMWT	925
OY	945	tcttatalaattttaaatacagcaagaagaaaaatagatctgtatlatlccaacata	1004
Db	924	MWTATATWTTTWTATRTARRARAATMTWTTTTTTTTTTTTTATATATTTTTTTTMYHHT	865
OY	1005	caatttagaacacacaatccagatataaaggtagatagtlactatccattatatt	1064
Db	864	TYTMTTHMHMMMMAMMMHMMMGGBCGKGGGTTTTTKTTTKTTTGCAATTTTA	805
OY	1065	tggtttaaaatataacatccaatcatatagtttllcatagaacaatacagaatlg	1124
Db	804	AATTTTTNRTDTWARAAADMDFTWMKTKMTHTTDMKKMKGMNNMMMVNGMTRTK	745
OY	1125	atttagtcccgaataaagaatattatagaattggtatagaanaacaatataataaaa	1184
Db	744	TYTKTKDGRRRMGMTDTBARAAAAAAAAGAPARAAARGAAGGGGGGRC	685
OY	1185	tcaaaataatacaggcagatggtatagaatagatcattatattaagaaga	1238
Db	684	GARGGAARDMKAANGAKMKTMTTRDARKRGAGGTTAATTTTTTTTMAAAR	631
RESULT	11		
LOCUS	CNS073Y8/C	DNA	GSS
DEFINITION	clone BA0AB02G11 of library BA0AB from strain CLIB 210 of Kluyveromyces lacticus genomic survey sequence.		
ACCESSION	AL428118		
VERSION	AL428118.1	GI:12211312	
KEYWORDS	GSS.		
SOURCE	Kluyveromyces lacticus.		
ORGANISM	Kluyveromyces lacticus.		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.		
AUTHORS	Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F., Duchateau-Nguyen, G., Imaeai, M., Marneisse, R., Montrocher, R., Robert, C., Temier, M., Winkler, P. and Wesolowski-Louvel, M.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lacticus		
JOURNAL MEDLINE REFERENCE	FEMS Lett. 487 (1), 66-70 (2000)		
AUTHORS	2 (bases 1 to 1161) Soucier, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Leplinge, A., Lorente, B., Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Potier, S., Saurin, P., Tekala, F., Toifano-Nioche, C., Wesolowski-Louvel, M., Winkler, P. and Weissenbach, J.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies		
JOURNAL MEDLINE REFERENCE	FEMS Lett. 487 (1), 3-12 (2000)		
AUTHORS	3 (bases 1 to 1161) Genoscope.		
TITLE	Direct Submission		

[illegible]

Query Match 4.48; Score 78.4; DB 13; Length 1101;  
Best Local Similarity 30.2%; Pied, No. 0.014;  
Matches 191; Conservative 158; Mismatches 276; Indels 7; Gaps

Oy 70 aaagaaaaataataaaatcattccgtgggaatatcacgagaataaaaataaacct 129  
||| ||| | : | | | : | : | : | : | : | : | : | : | : | : | :  
Db 472 AAAAAAATAAADGAAAAAAAAAAARRARATGARARTWRRARACTTAWTTGAIAAAAAAAA 531  
  
Oy 130 catcaacaggaaatacatgatgccttccttccacaacttaacgaataaatagtatttg 189  
: | | | | | : | | | : | : | : | : | : | : | : | : | : | : | :  
Db 532 AAATTTAAGCAAAAA--AAAWATATVKAAATTTTTTRRTBRACGTATTTTRDTRAkana 589  
  
Oy 190 gaatatgatccttatcgycggagatggtyttaaccttagaatcaaaaaaaaaataa 249  
Db 590 AAAAAATATTAGA AAAARGAAAAAARAAGTGTTTAGAACAAAAAATAADTA 649  
  
Oy 250 ataaataaaatgatagaanaatcaccagaacaattltagagaglatcaaaaagagaa 309  
: | | | | | : | | | | | : | : | : | : | : | : | : | : | : | :  
Db 650 AAKMMAAAAAGGGAAAAAATAAAAAAAMWTCTATAAAAAANAAMAARRGGAK 709  
  
Oy 310 caagttattcagcaggaatactaigtltcaaygaacagicaagtgatagagcaaacat 369  
Db 710 AGCGKRGARARAGRGGGRAMRAAAKTTRKKTKTRRAAGRARARAMGAAAAAAR 769  
  
Oy 370 atgattgatccaaccaatgytgtcaatttatatatgccatgctgcgcgaattaatgt 429  
Db 770 AADPAATKTGTAKKAAMRTAAKKKWCKTDAAAAKAARKTDYTTAKDKTKMGAAAA 829  
  
Oy 430 ttatatgtatccttagaccattlaaygaagaacaagaatlccattlaagaanaacgyc 489  
Db 830 TGAAGAAAGAAGRRGRKGDTFARARARAGAGDGMDKAAWMAAAATTTARMDTATAD 889  
  
Oy 490 ttgaatatgaaaacctatggtatgatgcttaatgaatlaagcgtatcaagaanaagta 549  
Db 890 DRRKAAWDKKRAAAPADDKDGRRAAGTWGMWRRAARTRKKAARCARAGAKKRRAK 949  
  
Oy 550 cceagatlaattaaatttattccttaggttggagaacccatatagatagataaanaa 609  
Db 950 GDAAKDDW-----KKMKDXMAAAMAANKGTGDMKKRARAKTCMGKABADKBMRRAADDT 1004  
  
Oy 610 gaaatgatcatgttggaataaataatgaagcttbttaatgatlttgattgaatgtgt 669  
Db 1005 RWRKRAMMTDWMAATATKDKTDKRAKCGKMAKTTTTRDRAAKAKADDTAATRTATAW 1064  
  
Oy 670 gatattgactgggaaccacgatggagttta 701  
Db 1065 KDADDKAKADAkakGCADkADkRADDMRDA 1096

RESULT 13 CNSOIU90/c 524 bp DNA GSS 12-MAY-2000  
LOCUS Tetradon nigroviridis genome survey sequence PUC-ORI end of clone  
DEFINITION 196C24 of library G from Tetradon nigroviridis, genomic survey  
sequence.  
VERSION ALI67541 GI:7805598  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorphae; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 524)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fitzmes,C., Fisher,C.,  
Bonneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished

[illegible]

	REFERENCE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
	AUTHORS	I (bases 1 to 876)			
	TITLE	Genoscope.			
	JOURNAL	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr  Web : www.genoscope.cns.fr) 			
	COMMENT	Determination of this BAC-end sequence was carried out as part of e collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> . location/Qualifiers			
	FEATURES	1..876			
	SOURCE	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1fb="RPCI-98" /clone="BACR19J14" /note="end : TPTT3"			
BASE COUNT	335 a	54 c	57 g	325 t	105 others
ORIGIN					
Query Match	Best Local Similarity	4.4%; Score 77.4;	DB 13;	Length 876;	
Matches	210;	Conservative	43;	Mismatches	265;
				Indels	3;
				Gaps	1;
OY	211 gatggtgtaactctagaattcaaaaaataaaatataaaaatatagtaganaa	270			
Db	358 KVAAGCGCAAAANAAAAACAANAANAANAANAANAANAANAANAANAANA	417			
OY	271 tcaccaaacgaattttagaggagcgcataaaaagaacaagtattatcacggatcc	330			
Db	418 AAAAATAATTAATAAAAATTAAANNAANAATAATTTATTTNAAWMTTATAATAAT	477			
OY	331 tatggttcacgagacgcgaagtgatagcgaaacacatgatctgacctcaaccocatg	390			
Db	478 TNATWTAWMATTTMAATATATATATATATATTTTTTAAAAATTAATAAAAAMWAHA	537			
OY	391 gtgtcaatttatatcatgccattgcgcgtcataaatatgltatalgatctagacca	450			
Db	538 TAATAAAAATTTATATAAAAAATTAAAMAATAATTTATATATATATATATATATAT	597			
OY	451 tttaaaggaaagaacattctctattaagaacaaacggcgttagatatgaaccttgt	510			
Db	598 TATTAATAATTTATATATAAATAAATAATTTATATATANAAATTTATATATATAT	657			
OY	511 atgatcctaatagaattagacgcatccaagaagaagtcgcccagatgaattattcttta	570			
Db	658 ATATATTTANNATPATATTTTMAATATTTMWTGAATATITNTATATATATATTTTATA	717			
OY	571 tccitagggtggaacacctalatagtatagatagaaaaagaattallatatlgytlaaa	630			
Db	718 NAAMATATATATANTAT	777			
OY	631 attctgaagctgttaatagtatcttgattagayggtttgatattctgcgtggaaacacat	690			
Db	778 WTMMNNNA--TTAATTTATWTATATATATATATATATATATATATATATATATATAT	834			
OY	691 gggaagtttacacctaaatgaatattcaaatta	731			
Db	835 ATTAT	875			

```

RESULT 15
BF259858/c 839 bp mRNA EST 23-FEB-2001
LOCUS HVSMEF0020H09f Hordeum vulgare seedling root EST library HVCDNA0007
DEFINITION (etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0020H09f, mRNA sequence.
ACCESSION BF259858
VERSION BF259858.2 GI:13120301
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM barley.
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 839)
Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Mair, D. and
Wood, T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11188971.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTACCTCCTCACTAAGG
High quality sequence stop: 825.
Location/Qualifiers
1. 839
/organism="Hordeum vulgare"
/db_xref="taxon:4513"
/clone="HVSMEF0020H09f"
/clone_lib="Hordeum vulgare seedling root EST library
HVCDNA0007 (etiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJCI21"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT 275 a 46 c 56 g 451 t 11 others
ORIGIN
Query Match 4.4%; Score 77.2; DB 11; Length 839;
Best Local Similarity 46.7%; Pred. No. 0.024;
Matches 263; Conservative 0; Mismatches 297; Indels 3; Gaps 1;
QY 75 aaaaaataataataatcatctgggaataaaggaataaataaataaactatca 134
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 AAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 728
QY 135 aacggaatacatgagcctttccacatcttaactgaataatgtaatttgagata 194
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 727 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATTTAATTAATT 668
QY 195 tggatcttatcgagatggtgtaactctagaaataaataaataaataaataa 254
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 667 TAATTTAATTAATAATTAATTAATAATAATAATAATAATAATAATAATAA 611
QY 255 taataatgataagaataccaccaagaacaatttagagagatataaataaagaacaagg 314
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATAATAATAATAATAA 551
QY 315 tatatagcaggtactactggttcagtcaaggtcaggtgagatgacaaacatagat 374
| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 550 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAAA 491
QY 375 tgattcaaacccaaatggtgcaatttatatcatgcatlgtcgcattatagltata 434
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 490 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTATATATATATATNNTT 431
QY 435 tgatgatacgaaccatttaaggaagaacaagatcctatlaagaacaacgcgttaga 494
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 430 TTTNTTTTATATATATATATATATATATATATATATATATATATATATATAT 371
QY 495 atatgaacctatgtagatgcttaatgaattgaacgtatcagaagaagtcacga 554
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 370 AATTTAAATTTTATATATATATATATATATATATATATATATATATATATAT 311
QY 555 tgaattatcttcatctccttaggtggaagaacctatagatagataagaagaagaat 614
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 TTTATTTTATTTAGTTATATATATATATATATATATATATATATATATATATAT 251
QY 615 tgattatgtagataaataatgca 637
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 TAGGATTTAATTAATAATAATAATAATAA 228

```

Search completed: March 21, 2002, 14:56:13  
Job time: 1904 sec

---

09/535066





OM of: US-09-579-383-2 to: A\_Geneseq\_1101.\* out\_format : pfs

Date: Mar 21, 2002 4:39 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODE=frame-n2p.model -DEV=xlp
-O/cq2.1/USPTO.spool/US09579383/runat.20032002.151238.29288/app-query.fasta.1.1849
-DB=A_Geneseq_1101 -QFMT=fasta -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORMEXT=0 -MILEN=0 -MAXLEN=200000000
-USER=US09579383.ecncl_1.87 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEX
-WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-579-383-2
Query length: 1764
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 104.710000
```

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AA84986 +		3130.00	5266.42	3.9e-286	587
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AA84985 +		635.00	1053.49	2.8e-51	378
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA55791 +		148.00	217.88	3.0e-05	1215
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA55791 +		148.00	217.88	3.0e-05	1215
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AA84984 +		148.00	217.88	3.0e-05	1215
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		144.00	206.85	8.3e-05	1817
/SID2/gcgdata/geneseq/geneseq/AA1995.DAT:AA84823 +		142.50	216.77	7.5e-05	561
/SID2/gcgdata/geneseq/geneseq/AA1996.DAT:AA84823 +		142.50	216.77	7.5e-05	561
/SID2/gcgdata/geneseq/geneseq/AA1997.DAT:AA84823 +		142.50	216.77	7.5e-05	561
/SID2/gcgdata/geneseq/geneseq/AA1998.DAT:AA84823 +		142.50	216.77	7.5e-05	561
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA84823 +		139.50	196.76	0.0002	2255
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		138.00	188.41	0.0004	3973
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		135.50	189.09	0.0006	2500
/SID2/gcgdata/geneseq/geneseq/AA1996.DAT:AA81825 +		130.00	191.04	0.0013	866
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		130.00	191.04	0.0013	866
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		126.50	190.76	0.0023	509
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		126.50	184.32	0.0029	934
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		126.50	176.36	0.0038	1979
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		125.50	188.41	0.0030	2485
/SID2/gcgdata/geneseq/geneseq/AA1996.DAT:AA81825 +		125.50	175.35	0.0050	1714
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		123.00	179.95	0.0059	807
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT:AA81825 +		122.50	161.78	0.0118	4134
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		121.00	161.86	0.0076	490
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		121.00	167.97	0.0121	1817
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		120.00	179.81	0.0095	507
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		120.00	164.22	0.0174	2206
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		119.50	164.43	0.0160	1987
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		119.00	169.09	0.0160	1188
/SID2/gcgdata/geneseq/geneseq/AA1997.DAT:AA81825 +		118.50	168.34	0.0178	1178
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT:AA81825 +		118.00	159.98	0.0256	2331
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		117.50	166.41	0.0223	1205
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		117.50	166.41	0.0223	1205
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		117.00	158.34	0.0318	2380
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		116.50	156.69	0.0360	2464
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		116.00	160.64	0.0345	1634
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		115.00	161.76	0.0389	1254
/SID2/gcgdata/geneseq/geneseq/AA1997.DAT:AA81825 +		115.00	161.76	0.0389	1254
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81825 +		114.50	171.30	0.0305	477
/SID2/gcgdata/geneseq/geneseq/AA1996.DAT:AA81825 +		114.50	154.18	0.0546	2366

/SID2/gcgdata/geneseq/geneseq/AA1998.DAT:AA86388 + 114.50 154.18 0.0546 2366  
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81816 + 114.00 169.04 0.0357 538  
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AA81824 + 113.50 159.30 0.0338 1245  
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT:AA86314 + 113.50 152.86 0.0669 2285  
/SID2/gcgdata/geneseq/geneseq/AA1998.DAT:AA84842 + 112.50 163.55 0.0546 711

#### seq\_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AA84986

seq\_documentation\_block:

ID AAB49986 standard; Protein: 587 AA.

AA849986;

13-MAR-2001 (first entry)

P. gallinaceum chitinase SEQ ID NO: 4.

Malaria; mosquito; chitinase; fungal disease; parasitic disease;

veterinary disease; arthropod pest.

Plasmodium gallinaceum.

MO200073488-A1.

07-DEC-2000.

26-MAY-2000; 2000WO-US14536.

28-MAY-1999; 99US-0136508.

03-FEB-2000; 2000US-0180051.

(TEXA) UNIV TEXAS SYSTEM.

Vinetz JW;

WPI: 2001-061553/07.

N-PSDB: AAC69669.

New nucleic acid encoding a Plasmodium species chitinase is useful for preventing transmission of malaria by mosquito feeding on subject that may harbor Plasmodium species organisms -

Claim 29; Page 125-127; 137pp: English.

The present invention provides the protein and coding sequences of the Plasmodium falciparum and P. gallinaceum chitinase enzymes. These

organisms are the cause of malaria in humans. The sequences are useful in the prevention and treatment of malaria, fungal diseases, parasitic

diseases and veterinary diseases, in preventing the transmission of malaria and in the control of arthropod pests in agriculture.

Sequence 587 AA;

#### alignment\_scores:

Quality: 3130.00 Length: 587  
Ratio: 5.332 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

#### alignment\_block:

US-09-579-383-2 x AAB49986 ..  
Align seg 1/1 to: AAB49986 from: 1 to: 587

1 ATGATTTTAAATATCAATATTTTAAATTAATTAATTAATTAATTAATTAATTC 50  
|||||  
1 Metasphenylsterileheulellelelalesterileleuryse 17  
51 TGCATTTCCGAACTTGAAGGAAAGAAATTAATTAATTAATTAATTCATTCG 100  
|||||  
17 TAlasnsenrAgtlrtrleulysgLylyAsnAsnllAsnAsnserleug 34  
101 GANTATATCGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG 150

```

|||||
34  lytlelearglulnslnlsnlystnhlsghlthrglullehsglu 50
151  TCTTTTCACATCTTAATCGAATATAGTATTTTGTAGAATATGATC 200
51  SerPheSerHisLeuLysSerAsnSerAsnProValGluTyrGlySe 67
201  TTATTCGGAGATGGGTGTAACCTAGAAATTCACAAAATAATAATAA 250
67  rTyrCysGlyAspGlyCysAsnSerArglleThrLysAsnAsnLysAsnI 84
251  TTAATTAATAATGATAGAAATCACCAGACAAATTTTAGAGAGATATAA 300
84  leAsnLysAsnAspArgLysSerProArgGlnlleLeuGlnGluTyrLys 100
301  AAAAGGAACAAGATATATATAGACAGATACTATGCTCATGGAACATCA 350
101  LysArgLysGlnGlyllellealaglyTyrTyrGlySerTTPAsnSerG 117
351  AGGTGATAGACCAAAACATATGATTGATTCAAACCCAAATGGTGCAATT 400
117  nGlyAspArgLysHisMetlleAspSerAsnProMetValSerlleL 134
401  TATATATGCTTTGCTCGCATATATATGATATGATGATCTAGACCA 450
134  euTyrlleAlaPheAlaArglleAsnMetLeuTyrAspValSerArgPro 150
451  TTTATGGAAGACAAAGATTCCTATTAAGAAAACAGGCTTAGATATAGA 500
151  PheAsnGlyArgGlnArgPheLeuLeuArgLysHisGlyLeuGluTyrG 167
501  AACCTATGATGATGCTTAATGAATTAAGACGATTCAGAAAAGTACGTC 550
167  uThrTyrGlyMetMetLeuAsnGlnlleArgArglleArgLysValArgp 184
551  CAGATGAATATATCTTTTATCCTAGTGGAGAAAACCTATATGTAAT 600
184  roAspValllelleLeuLeuSerLeuGlyGlyuThrTyrMetlleasp 200
601  ATAGAAAAGAAATTTGATGATGATTAATTAATTAAGCTTGTATAGCA 650
201  lleGluLysGlnlleAspTyrValAspLyslleLeuLysleuValAsnAs 217
651  TTTTGTATTTAGATGGGTAGATATTGACTGGGAACACATGGGAATTTT 700
217  pPheAspLeuAspGlyValAspIleAspTrpGlnProHisGlyLysPheT 234
701  ACACTTAATGAATTAATTTTCAATTTATATTAATTAATTAATTAAC 750
234  yTrsLeuAsnGlnLeuAsnProPheSerAsnTyrTyrilleLysleuLeAsn 250
751  TTGTTAAGAAAACATATTCGGAAGAAAAGTTAATTTTCAATTTGCTGTC 800
251  leuLeuArgLysThrilleProGlnGluLysleuLeuSerlleSerGlySe 267
801  ATCAAAATGCTGCATTTATCATGCGTTTCAAGAGTTGCATCTTCTGTAA 850
267  rSerAsnAlaAlaLeuSerCysValSerGlyValAlaSerPheCysLysA 284
851  ATGAGAATGTCATATACACTTAATTTTGTCTGAACAATTAAGAAACA 900
284  spoluGlnSerProTyrAsnThrLysPheLeuSerGlnGlnlleGluThr 300
901  AATAAGAAATTTACATAGGCGACGACGATGTTATCAGCAGAACTTTTAT 950
301  AsnLysGlnLeuHisArgAlaAlaAlaMetLeuSerAlaGlyThrPheIl 317
951  TAAATATTTTAAATACAGCAAGAGAAAATAGATCTTGATATTATTCAAA 1000
317  eaSnillePheAsnThrAlaLysGlnLyslleAspLeuValPheIleGlnT 334
1001  CATACAAATTTAGAAACTACAAATCCACATATTAATGATAGATATGACTTA 1050
|||||

```

```

334  hrTyrAsnLeuGlnuThrThrsAsnProAspIleMetValAspMetTyrLeu 350
1051  TCCCATTTATATTTTGGTTTAAATATATAACATCAATCATATTAAGTTT 1100
351  SerHisLeuTyrPheGlyLeuLysTyrAsnIleThrIlelleleuGlyPh 367
1101  TTCATTAACACATTAACAGAGGTGATTTAGTCCGAAAATTAAGAAATAT 1150
367  eSerLeuGlnHisAsnArgGlyGlyPheSerProGlnAsnLysGluLeuL 384
1151  TAGAATGGTAGGAAAACATACATGATTAATAATCAAAATTAATTAAGG 1200
384  euGlnLeuValGlyLysThrIleHisAspLysAsnGlnAsnAsnArg 400
1201  GCAGATGCTATAGGATATGCAATTTATTTTACAAATATTTGGAACATTTAA 1250
401  AlaAspGlylleGlylleThrPHisLeuPheMetLysGlnGlnLeuProth 417
1251  TGGATCATTCGATGTAGATATTTTCTTACAAATATTTGGAACATTTAA 1300
417  rGlySerPheAspValAspIlePheLeuThrsAsnIleTyrLysHisLeuA 434
1301  ATCCGAGTACCAAACTCCAAAGACCTACTATTAATGAAAACCTGAA 1350
434  snProGlnValGlnThrProLysAspLeuThrIleThnGlnAsnProGln 450
1351  GACTGTACACAAATGATGATATGATGTTCCAGAGCTGTTATTCACACAT 1400
451  AspCysSerThrIleAspGluTyrValProGlyLeuValIleProThrIl 467
1401  AGGATATATTTACAAACACAAATGATGCTATATGGAAGAAAACATGTTATT 1450
467  eGlylleTyrTyrLysHisAsnAspAlaIleTyrLysThrArgSerTyrS 484
1451  CAATTCATGCAACCTGGGTATAGACAGATATGAAATGGGACTGGTCAAGTA 1500
484  erlleHisAlaProGlyValAspArgTyrGlnTrpAspLeuValLysVal 500
1501  TGCTATGAAAATAATATGCGATGGGAAAAGCACCATTATTAACACTGA 1550
501  CysTyrGlnLyslleCysAspGlyLysAlaAlaHisTyrTyrAsnThrAs 517
1551  CTATTAAGAAAGCTCTATTTATTTATGGAAGGGAACCATATTTAATTA 1600
517  pTyrLysGlnSerSerllelleIleTyrLysGlyGluProTyrLeuIleL 534
1601  AATGGTGCAACAGAGACCTCCGAGATGACGACATGACATGATACACA 1650
534  yStrPTrpGlnGlnGlyProProGlnGlnAlaLeuGlnLysSerTyrThr 550
1651  AAACCTAGATGCATCCAAATGTCCAGGATAGAAAGATGAATAAATAATA 1700
551  LysLeuAspAlaSerLysCysProGlylleGlnuThrPAsnLysLysTyr 567
1701  TCCACATTAACACTAGAGTAGAGAGAACATATGAACAAGAGTGATTT 1750
567  rProHisLysProLeuGlnValGlnGlnGlnTyrGlnGlnGlnValAspL 584
1751  TACCATTACAA 1761
584  euProLeuGln 587

```

seq\_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB49985

seq\_documentation\_block:

ID AAB49985 standard; Protein; 378 AA.

XX AAB49985;

XX AC

XX DT

13-MAR-2001 (first entry)

XX DE

P. falciptarum chitinase SEQ ID NO: 3.

```

KM Malaria; mosquito; chitinase; fungal disease; parasitic disease;
KH veterinary disease; arthropod pest.
XX Plasmodium falciparum.
XX WO200073488-A1.
XX PD 07-DEC-2000.
XX PF 26-MAY-2000; 2000WO-US14536.
XX PR 28-MAY-1999; 99US-0136508.
XX RR 03-FEB-2000; 2000US-0180051.
XX PA (TEXA ) UNIV TEXAS SYSTEM.
PI Vinetz JM;
XX WPI: 2001-061553/07.
DR N-PSDB; AAC89668.
XX New nucleic acid encoding a Plasmodium species chitinase is useful for
PT preventing transmission of malaria by mosquito feeding on subject that
PR may harbor Plasmodium species organisms
XX Claim 29; Page 124-125; 137pp; English.
XX The present invention provides the protein and coding sequences of the
CC Plasmodium falciparum and P. gallinaceum chitinase enzymes. These
CC organisms are the cause of malaria in humans. The sequences are useful
CC in the prevention and treatment of malaria, fungal diseases, parasitic
CC diseases and veterinary diseases, in preventing the transmission of
CC malaria and in the control of arthropod pests in agriculture.
SO Sequence 378 AA:

alignment_scores:
    Quality: 635.00      Length: 345
    Ratio: 2.560         Gaps: 5
Percent Similarity: 71.884 Percent Identity: 37.391

alignment_block:
US-09-579-383-2 x AAB49985 ..

Align seg 1/1 to: AAB49985 from: 1 to: 378

265 AGAAAATACCAAGACAATTATTNAGAGGACTATAAAAAAGAAACAGG 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 ArgLysAsnProArgGluIleLeuThrPheLysGluSerLysGI 53
315 TATTTCACGAGTACTATGGTGTCATGCAGAACACCAAGCATAGAGCAA 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 YllelleglnelYTYrPrOserTrpValSerTYrAsnHisAsnLeu 70
365 AACATATGATTGATTCAACCACCAATGGCTCAATTTTATATTCGATT 414
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
70 ys.....AspleuasnProksnuLeusnvalYalHismetsSerphe 83
415 GGTCCGATTAATATGTTATATGATGTATCTGAACCATTTATAGGAAGA 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 AlAlysMetAsplSerTyraSpSerIlleGluseriLeValGlySerPr 100
465 AAGATTCGATTAAGAAAACACGCGCTTAGAATATGAACCTATGATGCA 514
::::::::::|||:|||||:|||||:|||||:|||||:|||||:
100 OleulePheLysSerLeuIlleGlyLeuGluTYrIlleGlyLeuAsnGluT 117
515 TGCTTAATGAATTAAGAGTATCTCAAGAAAGTAGCTCCAGATGTAATTAT 564
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 yzPheAsnAsplMeLerAsnLeuArGylsalArGrProAspIleIleMet 133
565 CTTTATCCTTAGGTGGAGAAACCTAT..ATGATAGATATAGAAAAAGA 611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

[illegible]

XX JP11313688-A.  
 PN  
 XX  
 PD 16-NOV-1999.  
 XX  
 PF 19-FEB-1999; 99JP-0041547.  
 XX  
 PR 20-FEB-1998; 98JP-0039285.  
 XX  
 PA (BEAB-) BE ABLE KK.  
 XX  
 DR MPI: 2000-057358/05.  
 DR N-PSDB: AA39342.  
 XX  
 PT A superheat-resistant chitinase -  
 PS  
 XX Claim 1: Page 7-12; 19pp; Japanese.  
 CC This represents a superheat-resistant chitinase. The protein can be  
 CC expressed by standard recombinant methodology. Chitinase hydrolyzes  
 CC chitin, and decomposes chitin to form chitobiose.  
 XX  
 SQ Sequence 1215 AA;

alignment\_scores:                   Quality: 148.00                   Length: 647  
                                   Ratio: 0.576                   Gaps: 28  
 Percent Similarity: 39.722       Percent Identity: 17.311

alignment\_block:  
 US-09-579-383-2 x AAY55791 ..

Align seg 1/1 to: AAY55791 from: 1 to: 1215

```

160 CATCTTAACGATAGTAATTTGTAGAAATGATGCTTATGCGG 209
   |||||
139 HisLeuysAlaasnThrThrTyrTyrGlyValValProvalLeuAl 155
210 AGATGGGTGTAACCTAGAAATTTACAAAATAATATAATATAATAA 259
   :|||||
155 aaspgly..... 157
260 ATGATAGAAATCACCAGACAATTT.....TTAGAG 291
   :||| ||||| :||| :|||
158 .serArgglySerProserAsnValLeuAlaIethrThrProleuAlu 173
292 GAGTATAAAAAGAAACAAGTATTATAGCAGATACATGCTGTCATG 341
   |||:
174 ProTyrArg.....ValIleValTyrTyrIleSerTr 184
342 GAACAGTCAAGGTGATAGACAAACAATATGATTCACAACCCATGG 391
   :||| :||| :||| :|||
184 pglArgTyrAla.....ArgLysPheTyrValSerAspIleProtrpg 199
392 TGTCATTTTATATATGATTCATTTGCGCATTAATATGTTATGANGTA 441
   :||| :||| :||| :|||
199 IulysValThrHisValAsnTyrAlaPheLeuAspLeuLysGluAspLy 215
442 TCTAGACCATTTAATGAGACAAAGAATTCCTATTAGAAACAACGCGTT 491
   :|||
216 ThrValAlaPhe..... 219
492 AGAATATGAACCTATGATGCTTAAT.....GAAATTAGAGTA 535
   |||||
220 ....TyrAspThrTyrAlaAspProleuAsnLeuGluAlaMetLysGluT 235
536 TCAGAAAAGTACGTCAGATGTAATTTCTTTATCCTTAGGTGAGAA 585
   :||| :||| :||| :|||
235 yrlsArgLysTyrProAlaValLysValLeuIleSerValGlyGlyTyr 251
586 ACC.....TATATGATGATATAGAAAAGCAAT 614
   ||| ||| |||

```

```

252 ThrLeuSerLysTyrPheSerValValAlaAlaAspProAlaLysArgL 268
615 TGATATATGCGATTAATAATATGAAAGCTGTAAATGATTTGATTTAGATG 664
   :|||
268 nArgPheAlaGluThrAlaIleGluIleLeuArgLysTyrAsnLeuSpG 285
665 GTGTAGATATTTGACTGGGAA.....CCACATGGGAGAGTTTTCACAC 705
   |||||
285 LylleAspIleAspTrpGluTyrProGlyGlyGlyMetAlaGlyAsn 301
706 TTAAATGAATTAATTTTCAAAATTTATATTTAAATTAATTAATCTGTT 755
   :||| :||| :||| :|||
302 TyrGluSerProAspAspGlyLysAsnPheValLeuLeuLeuLysAspLe 318
756 AAGAAAACATAATCCGGAAGAAAGTTAATTTCAATTTCTGTTTCACAA 805
   |||||
318 uArgGluAlaLeuAspLysAla..... 325
806 ATGCTGATTTATCATGCGTTTCAGAGTTGCATCTTCTGTAAAGATGAA 855
   :|||
326 .....AlaLysGluAsp 329
856 GAATCTCATATTAACACTAAATTTTGTCTGCACAAATAGAACAAATAA 905
   :|||
330 HisLysAspTyr..... 333
906 AGAATTACATAGGCGACAGCAGCATGTTATGACAGAACTTTTATTTATA 955
   :||| :||| :||| :|||
334 .....LeuLeuThrAlaAlaIleThrProAlaAsp 343
956 TTTTAAATACA.....GCAAAGAGAAATATAGAT 984
   :||| :||| :||| :|||
343 rovalLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359
985 CTGTGATTTATTCACAATAC..... 1005
   :||| :||| :||| :|||
360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376
1006 .....AATTGAACAATACAAATC 1024
   |||
376 rGlyHisLeuAlaProLeuTyrCysAspProAsnAlaProTyrThrAspG 393
1025 CAGATATA.....ATGTAGATATGTTACTA 1050
   :||| :||| :||| :|||
393 IuAsnValLysTyrHisPheCysValAsnTyrThrValGlnTyrTyrIle 409
1051 TCCCATTTATATTTGGTTAAATATACATCCATCATATATAGGTTT 1100
   :||| :||| :||| :|||
410 GlnHisVal.....ProAspLysThrLysIleThrValGlyLeuProPh 424
1101 TTCAATTAGAACATTAACAGAGGTGATTTAGTCCGAAATAATTAAGAAATAT 1150
   :||| :||| :||| :|||
424 e...TyrSerArgSerPheAlaAsnValProProGluAsnAsnGlyLeuT 440
1151 TAGAATTGGTAGAAAAACA..... 1170
   :|||
440 yrcInProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456
1171 .....ATACAGATTAATAATTAATAA 1193
   :||| :||| :||| :|||
457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnIleTr 473
1194 T..... 1194
   :|||
473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490
1194 ..... 1194
490 ePrSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506
1195 .....AATAGGCGAGATGATATAGGATATG 1220
   ||| ||| |||
507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523

```

```

1221 GCATTATTTATGAAGAACAATTACCACTGATCATTCATGATGATA 1270
      |||::: |||::: |||:::
523 pgluilethralaasparglysProglYThrAsnaphisProleu... 538
1271 TTTTCTTACAATATTTGGAACATTTAAT..... 1302
      |||::: |||::: |||:::
539 ....LeuAspThrValLeuGlnHisLeuGlyGlyProProAlaTTrp 553
1303 ...CCTGAAGTA.....CAAACTGCAAAAGACCTTACTAT 1334
      |||::: |||::: |||:::
554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrIva 570
1335 AACTGAA.....AACCTGAAGACT 1354
      |||::: |||::: |||:::
570 lProgluProThrProleuProProSerAsnGluThrThrProgluAspa 587
1355 GTACGACATA..... 1365
      |||::: |||::: |||:::
587 snGluThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603
1366 .....GATGAATATGTTCCA.....GACTCGTTATTCACACCAT 1400
      |||::: |||::: |||:::
604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620
1401 AGGATATATTTACAACAACATGAT.....GCTATATGCA 1435
      |||::: |||::: |||:::
620 rGlyAspPheValIysProGlySerLeuSerValIysValThrAspTrpG 637
1436 AAACCTAGATCTTATTCATTCATGCACCTGCTGATAGACAGATATGATG 1465
      |||::: |||::: |||:::
637 LysAnThrGluThrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653
1486 GACTTGTCGAAGTATGCTATGAAAAATATGCGATGGAAAGCAGCCCA 1535
      |||::: |||::: |||:::
654 ...ValValIysVal.....LysLeuIysAspGlySerSerValSe 666
1536 TTTATTTAACCTGACTATTA.....GAAAGCTCTATTATTATATGCA 1579
      |||::: |||::: |||:::
666 rSerPheThrSerAlaAsnIysAlaGluGlyGlyTyrValValPheT 683
1580 AAGGGAAACATATTTAATTAATGTTGGCAACAAGACCT 1620
683 hrProValSer.....TrpAsnArgGlyPro 691
seq_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB84784
seq_documentation_block:
ID AAB84784 standard; Protein; 1215 AA.
XX
AC AAB84784;
XX
DT 11-JUL-2001 (first entry)
XX
DE Pyrococcus heat resistant chitinase.
XX
KW Chitinase; heat resistant; disease.
XX
OS Pyrococcus sp. KOD-1.
XX
PN JP2001054381-A.
XX
PD 27-FEB-2001.
XX
PF 13-AUG-1999; 99JP-0229517.
XX
PR 13-AUG-1999; 99JP-0229517.
XX
PA (BEAB-) BE ABLE KK.
XX
WP1; 2001-310635/33.
DR N-PSDB; AAF87608.
XX

```

```

PT A highly heat-resistant chitinase
XX
XX Claim 3; Page 13-16; 19pp; Japanese.
XX
CC The present invention relates to a highly heat-resistant chitinase
CC maintaining at least 90% chitinase activity after treated at 100
CC degrees C for 1 minute. The chitinase can be used for the creation
CC of plants resistant to diseases. The present sequence is the
CC chitinase of the invention.
XX
SQ Sequence 1215 AA;

alignment_scores:
Quality: 148.00 Length: 647
Ratio: 0.576 Gaps: 28
Percent Similarity: 39.722 Percent Identity: 17.311

alignment_block:
us-09-579-383-2 x AAB84784 ..
Align seg 1/1 to: AAB84784 from: 1 to: 1215

160 CATCTTAATGCAATATAGTAATTTGTGACAAATATGATCTTATGCGG 209
      |||::: |||::: |||:::
139 HisLeuIysAlaAsnThrThrTyrTyrGlyValValProValLeuAl 155
210 AGATGGGTGAACCTCTGAATTTACAAAAAATAATAAATAATAATAA 259
      |||::: |||::: |||:::
155 aAspGly..... 157
260 ATGATAGAAATCACCAGACAAT.....TTAGAG 291
      |||::: |||::: |||:::
158 ..SerArgGlySerProSerAsnValLeuAlaIleThrThrProleuGlu 173
292 GAGTATAAAAAAGAAACAAGTATTATACAGATACCTGTTGCTCTG 341
      |||::: |||::: |||:::
174 ProTyrArg.....ValIleValIyrTyrIleSerTr 184
342 GAAACGTCGAAGGTATAGAGCAAAACATATGATGATTCAAACCAATGG 391
      |||::: |||::: |||:::
184 pGlyArgTyrAla.....ArgIysPheTyrValSerAspIleProTrpG 199
392 TGTCGAATTTATATATGCAATTCCTCGCATTAATATGTTATATGATGA 441
      |||::: |||::: |||:::
199 IulysValThrHisValAsnTyrAlaPheLeuAspLeuGlyAspGly 215
442 TCTAGACCATTTATGGAAGCAAAAGATTCTCTATTAGAAAAACGCGCT 491
      |||::: |||::: |||:::
216 ThrValAlaPhe..... 219
492 AGAATATGAACCTATGATGATGCTTAAT.....GAAATTAGACGTA 535
      |||::: |||::: |||:::
220 ...TyrAspThrTyrAlaAspProleuAsnLeuGluAlaMetIysGluT 235
536 TCAGAAAAAGTACGTCAGATGTAATTTATCTTTATCTTATGCTGAGAGAA 585
      |||::: |||::: |||:::
235 yrlYsArgIysTyrProAlaValIysValIleSerValGlyTyr 251
586 ACC.....TATATGATAGATATAGAAAAAAGAAAT 614
      |||::: |||::: |||:::
252 ThrLeuSerIlyTyrPheSerValValAlaAlaAspProAlaIysArgGly 268
615 TGATTATGTGATGAATAATATGAAGCTGTAAATGATTTGATTTAGATG 664
      |||::: |||::: |||:::
268 nArgPheAlaGluThrAlaIleGluIleLeuArgIysTyrAsnLeuAspG 285
665 GTGTAGATATGACTGGAA.....CCCATGGCAAGCTTTTACAAC 705
      |||::: |||::: |||:::
285 lYlIleAspIleAspTrpGluTyrProGlyGlyGlyMetAlaGlyAsn 301
706 TTAATGCAATTAATTTTCAAAATTAATTAATTAATTAATTAATTAAT 755
      |||::: |||::: |||:::

```

```

302 TyrGluSerProAspAspGlyLysAsnDheValLeuLeuLeuLysAspLe 318
756 AAGAAAACTATTCCGAGAAAGAAAGTAATTCAATTCTCGTCAACA 805
318 uArgGluAlaLeuAspLysAla..... 325
806 ATGCTGCATTATCATGCGTTTCAGAGTTGCATCTTCTGTAAGATGAA 855
326 .....AlaLysGluAsp 329
856 GAATGTCATATACACTAAATTTTGTGTGACAAATAGAAACAATATA 905
330 HisLysAspTyr..... 333
906 AGAATTCATAGGCGCAGCGAGTATACAGCAAGACTTTTATTATA 955
334 .....LeuLeuThrAlaIleThrProAlaAsp 343
956 TTTTAAATACA.....GCAAGAGCAAAATAGAT 984
343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTyrLeuAsp 359
985 CTGTATTTATTAACAACATAC..... 1005
360 SerIleAsnIleMetThrTyrAspTyrHisGlyAlaTrpGluThrIleTh 376
1006 .....AATTAGAACTACAAATC 1024
376 rGlyHisLeuAlaProLeuTyrCysAspProAsnIleProTyrThrAspG 393
1025 CAGATATA.....ATGCTAGATATGTACTTA 1050
393 LuAsnValLysTyrHisPheCysValAsnTyrThrValGlnTrpTyrIle 409
1051 TCCCATTTATTTTGGTTAAATATATACATACACATATTTAGTT 1100
410 GlnHisVal.....ProAspLysThrLysIleThrValGlyLeuProH 424
1101 TTCATTAGAACATACAGAGGTGATTTAGTCCGAAATAAAGATTAT 1150
424 e...TyrSerArgSerPheAlaAsnValProProGluAsnGlnLysLeuT 440
1151 TAGAATTGGTAGAAAAACA..... 1170
440 yrgLInProPheSerGlyThrProAlaGlyThrTrpGlyProAlaTyrGlu 456
1171 .....ATACATGATAAAATCAAAATATA 1193
457 ThrTyrGlyValMetAspTyrTrpAspValAlaGluLysAsnGlnSerSe 473
1194 T..... 1194
473 rGluTyrGluTyrHisTrpAspProIleAlaGlnValAlaTrpLeuTyrS 490
1194 ..... 1194
490 erProSerLysArgIlePheIleThrPheAspAspProArgAlaIleGly 506
1195 .....AATAGGCGAGATGATATAGGATATG 1220
507 IleLysValAspTyrMetLeuLysAsnGlyLeuGlyGlyValMetIleTr 523
1221 GCATTATTATTAAGAAAGCAATTACCACTGCATTCATGTAGATA 1270
523 pGluIleThrAlaAspArgLysProGlyThrAsnAspHisProLeu... 538
1271 TTTTCTTACAATATTTGCAACATTTAAAT..... 1302
539 .....LeuAspTrpValLeuGlnHisLeuGlyGlyLysProProAlaTrp 553
1303 ...CCTGCAAGTA.....CAAACTGCAAAAGCCTTACAT 1334
554 IleProAspThrTyrTyrIleGlySerAsnIleProSerAsnIleThrVal 570

```

```

1335 AACTGAA.....AACCTGAGACT 1354
570 lProGluProThrProLeuProProSerAsnGluThrThrProGluAspA 587
1355 GTAGCACAATA..... 1365
587 snGlnThrAsnProAsnProSerGlnGlyAsnGluThrAsnProAsnPro 603
1366 .....GATGATATGTGCA.....GGACTCGTATTCACACAT 1400
604 SerProGlyAsnGluThrThrProSerAspAsnGlnThrThrProSerTh 620
1401 AGGATATATTAACAACAAATGAT.....GCTATATGCA 1435
620 rGlyAspPheValLysProGlySerLeuSerValLysValThrAspTrpG 637
1436 AAAGTAGATCTTATTCATTCATGCACCTGGTGTAACAGATATGAATGC 1485
637 LysnThrGluTyrAspValThrLeuAsnLeuGlyGlyThrTyrAspTrp 653
1486 GACTTGTCGCAAGTATGCTATGAAAAAATATGCGATGGAGAAAGCGCCA 1535
654 ..ValValLysVal.....LysLeuLysAspGlySerSerValSe 666
1536 TTATTATTAACACTGACTATATAA.....GAAAGCTTATTTATATGCA 1579
666 rSerPheTrpSerAlaAsnLysAlaGluGluGlyGlyTyrValValPheT 683
1580 AAGGGAAACCATATTTAATTAAATGGTGCGCAACAGACCT 1620
683 hrProValSer.....TrpAsnArgGlyPro 691

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB18255
seq_documentation_block:
ID AAB18255 standard; Protein; 1817 AA.
XX
XX AAB18255;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:112.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX MPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; page 270-274; 577pp; English.
XX
XX The present invention describes proteins and their fragments (1) encoded
XX
CC

```

CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
CC vaccines against P. falciparum infection comprising (i) or (ii).  
CC (i) and (ii) are useful for the development of vaccines against  
CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (i), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (i) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

XX Sequence 1817 AA:

alignment\_scores:                      Quality: 144.00                      Length: 558  
   Ratio: 0.514                                      Gaps: 31  
Percent Similarity: 50.179              Percent Identity: 20.968

alignment\_block:

US-09-579-383-2 x AAB18255                      ..

Align seg 1/1 to: AAB18255 from: 1 to: 1817

```
64  ACCCTGAAGAGAAATAATATATATATTCATTCGGAGATATACGGGA 113
    ||| ::::::::::::::::::::|||:::
355 ThrLysGlnGlyLysHisAspMetAsnAsp.....Ty 365
114  AATATAAATAAATACTCATCAACGAAATACATGATCTTTTTCATC 163
    ||| ||||| ::::: ::::: ::::: :::::
365 rAsnMetAsnLysLysAsnAsnMetAspLLeasnlleThrIleAsnAsnA 382
164  TTAATGCAATATAGTAAATTTTGAGAAATNGATCTTATTCGCGAGAT 213
    ::::: ||||| ::::: ||| :::::
362 sNAsnAsnAsnAsnAsnAsnLysLysLLe.....Ty rAsnAspAsn 394
214  GGGCTAGTACTAGATTAACAAAAATAATAATAATAATAATAA.... 258
    ||| ::::: ::::: ::::: ||||| :::::
395 ThrLeuAsn.....ValTy rAsnAsnSerTy rAsnIleHisSerAsnHi 409
259  ....AATGATAGAAAATCACCAAGACAAATTTTAAAGAGATTAATAA 301
    ||||| ::::: ::::: ||||| :::::
409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGlnLysHisLeu 426
302  AAAGGAACAAGGTATTATAGCAGATACTATGCTCATGCAACAGTCAA 351
    || ::::: ::::: ::::: ||| :::::
426 ys.....MetLeuCysAspAsnPh.....PheAsnLeuGln 436
352  GGTGATAGAGCAAAACATATGATGATTCACCAACCATGATGTCATTTT 401
    ::::: ::::: ||| :::::
437 GluPheTy rSerSerAsnIleIleIleAsnAsnMetAspIleGluTy rTh 453
402  AATATATTCATTTGCTGCATTAATATATGTTATATGATGATATCAGCAT 451
    ||| ||| ::::: ||||| :::::
453 Ty rAspTy rPhe.....IleLeuTy rGlu.....LysCysP 464
452  TTAATGAGAAGCAAGATTCCTATTAAGAAAAACAGCGCTAGAAATATGA 501
    || ::::: ||| :::::
464 HeLeuProIleGluIleVal.....HisValAsnTy rMet 476
502  ACCATATGATGATGCTTAATGAATATGACGATATCAAAAAAGTACGCC 551
    ||| ::::: ||||| :::::
477 LysTy r...LeuTy rLysAsnAsnGlnLysLysAsnLysIleArgLys 492
```

```
552  AGATGAATATTTATCTTTATCTTAGGTGAGAACCTATATGATAGATA 601
    ::||| ||||| ::::: ||| |||||
492 sPheLeuIleThrLeuLeuGln.....Ty rSerArgAspI 504
602  TAGAAAAAGCAATTCATATATGTCGATTAATAATTTG...AAGCTTGTAAT 648
    ||::: ::::: ::::: ||||| ||| |||
504 LeasnPheHisPhePheIlePheAsnLeuIleLeuTy rLysCysLysAsn 520
649  GATTTCGATTTAGATGATGATAGATTTAGCTGCGAACAACATGGAAGAT 698
    ::||| ::::: ::::: :::::
521 GluPheProCysSerIlePheGluLeuHisIleSerGlnTy rLeuTy rPh 537
699  TTAC...AATTAATGAATTAATTAATTTTCAATATATATAT..... 738
    ||::: ::::: ||||| :::::
537 ePheValLysLeuAsnGlnLeuAsnIleLysAspLalTy rIleTy rTy rP 554
739  ....AATTAATTAATCTGTTAAGAAAAACT 765
    ::||| ::::: ||||| :::::
554 HeAsnAsnPheLysTy rGlnAspMetIleIleTy rPheSerArgLysAla 570
766  ATT...CCGAGAGAAAGTTAATTCATTTCTGTCATCAAAATGCTGC 812
    ::||| ||| ::::: :::::
571 PheTy rProIlePheGln.....ThrAsnValGlnGlnLysLysGlnTh 585
813  ATTATCATGCGTTTTCAGAGATTCATCTTCTGTAAGATGAAGATCTTC 862
    ::||| :::::
585 rLeuSerTy rIle..... 589
863  CATATACACCTAAATTTTGTCTGACAAATAGAACAAATTAAGAAATTA 912
    ||||| ||| ::::: ||||| :::::
590  ..Ty rAsnAspLys.....IleLysLysAsnLysLys... 599
913  CATAGGCGACGACGCGATGTTATCAGCAGAACTTTATTAAT..... 954
    ::::: :::::
600 AsnAsnSerTy rTy rGluMetAsnAsnAsnThrTy rMetAsnGlnHisG 616
955  ....ATTTTAATACAGCAAGAGAAATATAG 982
616  YTy rThrAspIleGlnAsnGlnLysArgLeuAsnLysLysAsnLysArgLeu 633
983  ATCTGTATTTATTCACAAATACATTTAGAAATACCAATACCAATATTA 1032
    ::||| :::::
633 sNVal.....ArgGlyArgTh rAsnThrLeuAspAspIle 644
1033  ATGATAGATAGTACTATCCCATTTATATTTGGTTTAAATATATACAT 1082
    ::||| :::::
645 IleValSer.....AspHisGlyAsnSerTy rAspLysTy rAsnTh 658
1083  CACAATCATATAGTTTTTCATTGAAACATPACAGAGGTGATTTAGTC 1132
    :::::
658 rSer.....LysHisAsnArg..... 663
1133  CCGAAATATGAAGATTTATAGAAATGGTAGAAAAACAATACATGATATA 1182
    ::||| :::::
664  ....ArgLysAsnHisIleAsnGlnMetLysLysLysGlnAsnLys 678
1183  AATCAAAATTAATATAGGCGAGATGTTATAGGATATGCGATTTATTTAT 1232
    ::||| :::::
679 LysLysAsnThrLeuPheValAspLysAspMetGlnGlyIleGlyLys 695
1233  GAAAGACAAATTAACCACTGCATTCGATGATAGATATTTTCTTTCANA 1282
    ::||| :::::
695 sGlnLysGlnLysGlnAsnLysAsnMetAsnAsnIlePheTy rAsnA 712
1283  AT.....ATTGGAAACATTTAAATCCTGAA 1308
    || :::::
712 sNserTy rSerAsnIleAsnAsnSerSerTy rSerAsnIleAsnAsn 728
1309  GTACAAACTCCAAAAGACCTTACTATTAACGTGAAACCTCGAAGACTGTAG 1358
    :::::
729 IleTy rSerValAspAsnMetThrSerValAsnAsnThrLysTy rValSe 745
```





```

236 1nlvsa1gl1n1vsg1y1al1vhr1alatr1pa1p1a1p1roty1v1y1vsg1y1asn 252
517 CTTAATGAAATTAGACGATATACAGAAAGTACGTCACATATGTAATATATCT 566
      :::::::::::::::::::::::::::::::::::::::::::::::::::: |||||
253 p1vsg1y1n1l1v1m1e1t1a1l1e1u1d1y1s1g1n1a1h1s1p1ro1s1p1e1u1y1s1l1e1e 269
567 TTTATCTTAGGTGAGGAACCC.....TATATGATAG 598
      | ||||| ||| |
269 u1b1o1s1r1l1e1g1y1t1p1h1r1l1e1u1s1e1r1a1s1p1r1o1p1h1e1p1h1e1m1e1c1y1a 286
599 ATATGAAAAAGAAATTCATTATGCTATGCTAAATATTGAAAGCTGTTAAT 648
      || |::: ::::: |||
286 s1p1y1s1v1a1l1y1s1a1r1g1s1p1a1r1p1h1e1v1a1g1y1s1e1r1v1a1l1y1s1g1u1p1h1e1l1e1g1n 302
649 GATTTTGAT...TTAGATGCGTGTGATGATATGACTGGGAA...CCACATG 692
      :::::::::::::::::::::::::::::::::::::::::::::::::::: ||
303 t1h1r1t1p1y1s1p1h1e1p1e1s1p1g1y1v1a1s1p1l1e1s1p1r1p1o1h1e1p1o1g1y1 319
693 GAAGTTTATC.....AACTTAAATGAAATTAATTTTCAAAATTAATATA 736
      |||| |::: |||||::: |||
319 y1vsg1y1a1l1a1s1n1p1ro1s1n1l1e1u1g1y1s1e1r1p1ro1g1n1a1s1p1g1y1c1u1t1h1r1t1y1v 336
737 TTTAATTTAATTAAGTGTGTAAGAAAACATATCCGGAGAAAAGTTAAT 786
      :: |||::: |||||
336 a1l1e1u1l1e1m1e1t1y1s1g1l1e1u1n1r1g1a1l1a1m1e1t1e1u1s1p1g1n..... 348
787 TCAATTTCTGGTTCATCAAAATGCTGATTCATTCATGCGCTTCAGAGTGC 836
      |||||::: |||||
349 .....l1e1s1e1r1a1g1l1u1r1h1g1y1..... 354
837 ATCTTTCGTAAAGATGAGAAATTCATATTAACACTAAATTT...TTGT 883
      |||||::: |||||
355 .....A1r1g1y1s1t1y1r1g1l1e1u1t 360
884 CTGAACAATAGAAACAATAATGAATATACATAGAGGACAGCAGATGTA 933
      ::::: |||::: |||:::
360 h1s1e1r1a1l1l1e1s1e1r1a1g1l1y1s1p1s1y1l1e1s1p1y1s1v1a1l1a..... 374
934 TCACGAGGAACTTTAATTAATTTTAAATACAGCAAAAGAGAGAAATAGA 983
      ::::: |||::: |||:::
375 .....T1y1a1s1n1v1a1l1a1g1l1a1s1n1s1e1r1m1e1t1a1s 383
984 TCTGTATTATTCAAACATACAT 1008
      | ::|||::: |||:::
383 p1h1s1t1l1e1p1h1e1l1e1u1e1s1e1r1t1y1r1a1s1p 391

seq_name: /SID22/gcgdata/geneseq/geneseq/A11996.DAT:AAW09402
seq_documentation_block:
ID AAW09402 standard; Protein; 561 AA.
XX
XX AAW09402:
XX AC
XX DT
XX 19-FEB-1997 (first entry)
XX DE
XX Serratia marcescens Chitinase A.
XX
XX Chitinase; Ch1a; chitin degradation; transgenic plant; tomato;
XX cold resistance; tolerance; frost damage; reducing sugar content;
XX increase; ATCC 9901; ds.
XX
XX Serratia marcescens QMB 1466.
XX OS
XX
XX Key Location/Qualifiers
XX FT 1..23
XX FT /label= signal_peptide
XX FT /note= "predicted cleavage site is between residues
XX FT 23 and 24"
XX FT Protein
XX FT 24..561
XX FT /label= Chitinase_A
XX FT /note= "predicted mature sequence"
XX
XX US5554521-A.

```

XX	
PD	10-SEP-1996.
XX	
PF	26-MAR-1984;      84US-0593691.
XX	
PR	19-DEC-1994;      94US-0356901.
PR	26-MAR-1984;      84US-0593691.
PR	18-JUL-1986;      86US-0888033.
PR	09-JUL-1990;      90US-0505253.
PR	14-AUG-1992;      92US-0930970.
PA	(DNAP ) DNA PLANT TECHNOLOGY CORP.
PI	Jones JDG, Sushow TV;
PI	
DR	WPI; 1996-424655/42.
DR	N-PDB; AAT9479.
XX	
PT	Recombinant plant with increased resistance to cold - and increased PT reducing sugar content, contains DNA encoding chitinase, partic. to PT protect tomatoes against frost damage.
PS	Example 3; Fig 1A-C; 25pp; English.
CC	
CC	The present sequence is the Chitinase A enzyme encoded by the 2.3 kb CC Insert of plasmid pCHIT1251 which was isolated from genomic DNA of CC <i>Serratia marcescens</i> OMB 1466 (ATCC 9901). The chlA gene can be used for transforming plants, esp. tomatoes, to give increased resistance to frost damage and to produce sweeter fruit.
Sequence	561 AA;

[illegible]

123 *rasp1a1*trg1u1levalvala1laasprthar9sern1leu1pro 140  
268 .... AATACCAACCAACAAATTTTACAGACAGATATAAAAAAGAACAA 312  
140 ro1eu1ys1u1pro.....Leu1leu1l1u1s1n1s1p1ro1u1tr1y1s1u1 154  
313 GGT.....ATTATACAGAGATACATAGCTTCAATGCTTCAAGCAACAG 353  
155 Asnserg1LysvalVal1g1ser1u1tr1pheval1G1ut1re1l1val1u1tr1 171  
354 TGATAGACCAAAACATATGATTTGATTCACAAACCAATGCTGTCAATTTAT 403  
171 Y.....Argasnp1r1eth1Val1Asp1ys1le1Pro1a1Gln1Asn1Leu1thn 186  
404 ATATGCAATTTGCTGTGCATTT.....AATATGTTATATGAT 438  
186 1s1leu1en1u1tr1c1y1phe1le1pro1le1c1ys1G1u1Lysn1u1le1Asn1sp 202  
439 .....GTATACACCAATTTAA 455  
203 Ser1eu1ys1G1u1le1G1u1Lys1er1phe1Gln1a1le1Gln1a1r1g1ser1c1ys1 219  
456 TGAGAGCAACAAAGATTCCTATTAAGAAACAC.....GGCTTAG 493  
219 ngl1y1r1g1c1u1as1r1phe1lys1le1ser1le1h1s1as1p1ro1phe1la1a1leu1 236  
494 AATATGAACCTATGATGATGATG.....TATATGATAG 598  
236 1u1Lys1a1G1u1Lys1e1Lys1Val1Th1ra1a1tr1p1As1p1ro1u1tr1y1s1u1Asn 252  
517 CTATATGAATTAAGACGATATCGAAACAGACCTCCGATGATATTTCT 566  
253 Phe1g1Lys1u1Leu1Me1ta1Leu1u1ys1Gln1a1h1s1p1ro1Lys1u1le1u1 269  
567 TTTATCCTTAGGTGAGAAACC.....TATATGATAG 598  
269 u1p1ro1ser1le1c1y1G1u1tr1p1th1Leu1Ser1as1p1ro1phe1r1h1em1ec1L1A 286  
599 ATATAGAAAAGAAATTCATTTATGTGGATTAATAATATGAACCTTTGTAAT 648  
286 s1p1Lys1Val1Lys1a1r1g1As1p1a1r1g1Phe1Val1G1y1Ser1Val1Lys1u1Phe1u1Gln 302  
649 GATTTTGAT...TTAGAGTGCTAGATATTTGACTGGAA...CCAACATAG 692  
303 Thr1tr1p1Lys1Phe1h1as1p1c1Y1a1Lys1p1le1as1p1r1e1G1u1Phe1r1o1c1y1G1 319  
693 GAAGTTTAC.....AACTTAATGAATTAATTTTCAAAATTTATATA 736  
319 Y1Lys1G1Y1a1La1snp1ro1Asn1Leu1G1y1Ser1Pro1Gln1s1p1G1u1Th1r1Y1V 336  
737 TTAATTAATTAATTAACCTTTTAAGAAAACATATTCGAGAAAGAAAGTTAAT 786  
336 a1le1Leu1e1Me1L1ys1G1u1Leu1a1r1G1a1Me1t1Leu1Asp1Gln..... 348  
87 TCAATTTCTGGTTCAATCAATGCTGCATCATTCATCATCTAAATTT...TTC 883  
349 .....Leu1Ser1a1a1G1u1Th1G1Y..... 354  
837 ATCTTTCTGTAAAGATGAAGATCTCCATATTAACCTAAATTT...TTC 883  
355 .....Arg1ys1Tyr1G1u1Leu1T 360  
884 CTGAACAATAGAAACAATTAAGAATTAACAATAGGCGACGACGCGATGTTA 933  
360 h1se1r1a1a1le1Se1r1a1G1y1Lys1as1p1Lys1le1As1p1Lys1Val1a1a1..... 374  
934 TCAAGCAACACTTTTATTAATATTTTAAATACAGCAACAAAGAAATAAG 983  
375 .....Tyr1as1n1Val1a1Gln1Asn1Ser1Met1as 383  
984 TCTTGTATTTATTCGAACATTCGAAT 1008  
883 r1h1s1le1Phe1Leu1Met1Ser1u1Y1Asp 391

```

44  spGlnAlaIAlaThr.....AlaTyrAsnAsnLeuValIysVal 56
166  AATCGAATAATAGTAATTTGTAGAAATATGACCTTATTCGAGAT.. 213
    |||::: |||::: |||::: |||::: |||::: |||:::
57  LysAsnAlaIAlaaspValSerValSerTrrpAsnLeuTrrpAsnGlyAspAl 73
214  .GGGTACTCTGTAGAAATTACAAAATTAATAA..... 246
    |||::: |||::: |||::: |||::: |||::: |||:::
73  aglyThrGlyProIysIleLeuAsnGlyLysGlnAlaTrrpSerGlyP 90
247  .....AATATAATAATAAATCAT 264
90  roSerThrGlySerSerGlyThrAlaAsnPhelyAlAsnLysGlyGly 106
265  AGA..... 267
107  ArgTyrGlnMetGlnValAlaIAlaLeuCysAsnAlaaspGlyCysThrAlase 123
267  ..... 267
123  rAspAlaThrGluIleValValAlaAspThrAspGlyArgHisLeuProp 140
268  ....AATCACCAGAACAAATTTTAGAGAGTATTAATAAAGAACAA 312
140  rOleuLysGluPro.....LeuLeuGluLysAsnLysProTyrLysGln 154
313  GGT.....ATTATGACAGATACTACTGTTCAATGACAGACACTCAAG 353
155  AsnSerGlyLysValValGlySerTyrPheValGluTrrpGlyValTyrGln 171
354  TGAATAGACAAACATATGATTCACAAACCAAGGTGTCATTTTAT 403
171  Y.....ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186
404  ATATTGCATTGCTGCATTT.....AATATCTTATATCAT 438
186  ILeuLeuTyrGlyPheIleProIleCysGlyLysGlnLysAsnAsp 202
439  .....GTATCTAGACCATTTAA 455
203  SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnAlaGserCysGln 219
456  TGAAGACAAAGATTCTTATTAGAAAAC.....GCCTTAG 493
219  nGlyArgGluAspPheLysIleSerIleHisAspProPheAlaIAlaLeuG 236
494  AATATGAACCTATGTTATGATG..... 516
236  InLysAlaGlnLysGlyValThrAlaTrrpAspAspProTyrLysGlyAsn 252
517  CTTATGAATAATTAGACGTATCAGAAAGTACGTCAGATGTAATTATCT 566
253  PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269
567  TTTATCTTACGTGAGAACCC.....TATATGATAG 598
269  uProSerIleGlyTrrpThrLeuSerAspProPhePhePheMetGlyA 286
599  ATATAGAAAAGAAATGATTATGATTAATAATATGAACTTGTTAT 648
286  sPLysValLysArgAspArgPheValGlySerValLysGlnPheLeuGln 302
649  GATTTGAT...TTAGATGTTGATATTTAGCTGAGGAA...CCACATGG 692
303  ThrTrrpLysPhePheAspGlyValAspLysIleAspTrrpGluPheProGlyGln 319
693  GAAGTTTAC.....AATTAATGAATTAATTTTCAATATATATA 736
319  yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyLysIleuThrYrv 336
737  TTAATTTAATTAACCTGTTAGAAAACATATTCGGAAGAAAGTAAAT 786
336  alLeuLeuMetLysGlnLeuArgAlaMetLeuAspGln..... 348

```

```

787  TCAATTTCTGGTTCATCAAAATGCTGATTTATCATGCGTTTCAGAGATTGC 836
349  ..... |||::: |||::: |||::: |||::: |||::: |||:::
837  ATCTTTCTGTAAGATGAGAAATCTCCATATACACTAAATTT..TTGT 883
355  .....ArgLysTyrGlnLeuT 360
884  CTGAACAATAGAAACAATAAAGAAATTTACATAGGCGCAGCAGCATGTTA 933
360  hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374
934  TCACGACGAACTTTATTAATATTTTATATACAGCAAGAGAAATAGA 983
375  .....TyrAsnValAlaGlnAsnSerMetAs 383
984  TCTGTATTTATTCAAACATACAT 1008
383  rHisIlePheLeuMetSerTyrAsp 391

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT.AAM60179
seq_documentation_block:
ID  AAM60179 standard; Protein; 561 AA.
AC  AAM60179;
XX
DT  03-SEP-1998 (first entry)
XX
DE  Serratia marcescens chitinase A sequence.
KW  Chitinase A; pathogen; plant protection; Serratia marcescens; E. coli;
KW  pseudomonas; Erwina; frost damage resistance; plant pathogen; fungus;
KW  nematode.
OS  Serratia marcescens.
XX
FH  Key Location/Qualifiers
FT  Peptide 1..23
FT  Cleavage-site 23..24 /note= "putative signal peptide"
FT  Protein /note= "predicted cleavage site"
FT  24..561 /note= "mature protein"
XX
PN  US5776448-A.
XX
PD  07-JUL-1998.
XX
PF  01-AUG-1996; 96US-0693835.
XX
PR  19-DEC-1994; 94US-0358901.
PR  26-MAR-1984; 84US-0593691.
PR  18-JUL-1986; 86US-0888033.
PR  09-JUL-1990; 90US-0550253.
PR  14-AUG-1992; 92US-0930970.
PR  01-AUG-1996; 96US-0693835.
XX
PA  (DNAP ) DNA PLANT TECHNOLOGY CORP.
XX
PI  Jones JDG, Suslow TV;
XX
DR  MPI; 1998-397941/34.
XX
DR  N-PSDB; AAV35601.
XX
PT  Use of heterologous chitinase nucleic acid - for transforming
PT  bacteria or plants, for increasing plant resistance to plant
PT  pathogens, increase chilling resistance and increasing sweetness.
XX
PS  Example 3; Fig 1A-C; 25pp; English.
XX
CC  This represents a Serratia marcescens chitinase A. The invention

```

CC provides a method for the production of a bacterial cell capable of  
 CC producing chitinase which comprises transforming the bacterial cells  
 CC with a nucleic acid encoding the chitinase, the nucleic acid being  
 CC isolated from a heterologous source, and whereby the bacterial cell  
 CC is capable of expressing the nucleic acid. The bacterial cell is selected  
 CC from *Pseudomonas* sp., *E. coli* or *Erwinia* sp. The products can be used to  
 CC enhance plant growth by biological control of plant pathogens such as  
 CC fungi, nematodes, insects and disease agents. Plants transformed with  
 CC the chitinase DNA can also have resistance to frost (freezing) damage  
 CC or chilling damage, increased levels of reducing sugars or sweetness  
 CC in fruits or plants and enhanced post-harvest storage life. The  
 CC products can also be used for the production of chitinase for use as an  
 CC antibiotic.

XX Sequence 561 AA:

alignment\_scores:      Quality: 142.50      Length: 425  
                           Ratio: 0.720      Gaps: 20  
                           Percent Simlarity: 46.588      Percent Identity: 20.471

alignment\_block:

US-09-579-383-2 x AAW60179 ..

Align seg 1/1 to: AAW60179 from: 1 to: 561

```

25  TTAATTATAGTATCCATCTGTATTCGCAAAATTCACAGACCTTGAAAG 74
    |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11  LeuLeuIleGlySerThrLeuCySerAlaIleGlnAlaIleAlaProG1 27
75  AAAAAATAATAATA.....AATAATCTATGGGAATAATACGGCAA 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27  YLpSPProThrIleAlaTrpGlyAsnThrLYSpheAlaIleValGluVal 44
116 ATAAAAATAAACATCATCAACGCAAAATACATGAGCTTTTTCACATCTT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44  spGlnAlaIleThr.....AlaLYSpheAsnLeuValLYSpVal 56
166 AAATCGAATAATAGTAAATTTGTAGAAATANGAGATCTTATTCGGAGAT 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57  LYSpAsnAlaIleAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73
214 .GGGTGTAAGCTAGAAATTAACAAAAATATAATAA..... 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73  agLYThnGlyProLYSileLeuLeuAsnGlyLYSgluAlaIleTrpSerGlyP 90
247 .....AATATAAATAAAATGAT 264
90  roSerThrGlySerSerGlyThrAlaAsnPhelyValAsnLYSGlyGly 106
265 AGA..... 267
107 ArgTYGlnMetGlnValAlaLeuCyAsnAlaAspGlyCysThrAlase 123
267 ..... 267
123 rasPalatrhGluIleValAlaAspThrAspGlySerHisLeuProP 140
268 .....AAATCCACCAAGCAAAATTTTAGAGAGATATAAAAAAGCAACAA 312
140 roLeuLYSGluPro.....LeuLeuGluLYSAsnLYSpProTYrLYSGln 154
313 GGT.....ATTATAGCAGATACTATGTTTCATGACGAACAGTCAAG 353
155 AsnSerGlyLYSValValGlySerTYrPheValGluTrpGlyValTYrG1 171
354 TGATAGACAAACATATGATGATTCAAACCCAAATGGTGTCAATTTAT 403
171 Y.....ArgAsnPhetThrValAspLYSileProAlaGlnAsnLeuThrH 186
404 ATATTCATTTGCTGCATTT.....AATANGTATATATGAT 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

186  ISeuLeuTYrGlyPheIleProIleCySGlyGlyAsnGlyIleAsnAsp 202
439 .....GTATCTAGACCATTTAA 455
203 SerLeuLYSGluIleGlySerPheGlnAlaLeuGlnArgSerCYSG1 219
456 TGAAGACAAAGATTCCTATTAGAAAACAC.....GGCTTAG 493
219 nGlyArgGlnAspPheLYSileSerIleHisAspProPheAlaIleLeuG 236
494 AATATGAACCTATGCTATGATG..... 516
236 InLYSAlaGlnLYSGlyValIleThrAlaTrpAspAspProTYrLYSGlyAsn 252
517 CTATATGAAATTTAGACGTATCAGAAAGTACGTCAGATGTAATTTATCT 566
253 PheGlyGlnLeuMetAlaLeuLYSGlnAlaHisProAspLeuLYSileLe 269
567 TTATTCCTTAGGTGAGAAACC.....TATATGATGAT 598
269 uProSerIleGlyLYrTrpThrLeuSerAspProPhePhePheMetG1YA 286
599 ATATAGAAAAAGAAATGATGATGATGATGATAAATATGAAAGCTTGTAAT 648
286 sPLYSValLYSArgAspArgPheValGlySerValLYSGluPheLeuGln 302
649 GATTTTGAT...TTGATGGTGTAGATATTGACTGGGA...CCACATGG 692
303 ThrTrpLYSPhePheAspGlyValAspIleAspTrpGluPheProGlyG1 319
693 GAAGTTTAC.....AACTTAATGCAATTAATTTTTCAAATTTATATATA 736
319 YLYSGlyValAsnProAsnLeuGlySerProGlnAspGlyGluThrTYrIV 336
737 TTAATTTATTAATCTGTTTAGAAAAAATACTATTCGGAAGAAAGTTATAT 786
336 alLeuLeuMetLYSGluLeuArgAlaMetLeuAspGln..... 348
787 TCAATTTCTGTTTCATCAAAATGCTGATTTATCATGCGTTTCAGAGTTGC 836
349 .....LeuSerAlaIleuThrGly..... 354
837 ATCTTTCTGTAAAGATGAGATCTCCATATATACACTAAATTT...TTGT 883
355 .....ArgLYSTYrGluLeuTYr 360
884 CTGAACAATAGAACAAATAAAGCAATTTACATGAGCGACGACGAGATGTA 933
360 hrSerAlaIleSerAlaGlyLYSAspLYSileAspLYSValAla..... 374
934 TCAGCAGGAACCTTTATTTAATTTTAAATACAGCAAGAGAGAAATAGA 983
375 .....TYrAsnValAlaGlnAsnSerMetAs 383
984 TCTTGTATTATTTCAAACATACAT 1008
383 pHisIlePheLeuMetSerTYrAsp 391

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAB18180

seq\_documentation\_block:

ID AAB18180 standard; Protein; 2295 AA.

XX AAB18180;

DT 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:37.

DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

KW antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

OS







CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

xx  
SQ Sequence 3973 AA:

# alignment\_scores:

Quality: 138.00 Length: 675  
Ratio: 0.471 Gaps: 34  
Percent Similarity: 43.407 Percent Identity: 20.889

## alignment\_block:

US-09-579-383-2 x AAB18253 ..

Align seg 1/1 to: AAB18253 from: 1 to: 3973

```

1  ATGATTTTAAATATCATATTTTAAATATATATATCCATCTGTAT.. 48
   |||
2770  ILeasnlleuValasnLeuphelelleasnllysargaspTyrAs 2786
   |||
49  ....TCGCAAAATTCAGAACCTTGAAAGAAAAATATATATAATA 91
   |||
2786  pasgclutHrcysmetSerAsnlelleasnasnspasnslsYsa 2803
   |||
92  ATTCAATGGCAATATATACGGAAAAATAAAATAACTCATCAACGAA 141
   |||
2803  snlYsaAsnleuile...GluasnlysngluileTyrAsnThrAsn 2818
   |||
142  ATACATAGAGCTTTTTCACATCTTAA.....TCGATATATAGTAA 182
   |||
2819  Ile...LysSerleuYsaAsnspLysglutYrIleasnasnHisSerAs 2834
   |||
183  TTTTGTAATATGATGATTTATTCGAGATGGGT..... 219
   |||
2834  nTyrAlaMetHe....TyrCysAspLeuPheCysAspAspPheHei 2849
   |||
220  ....AAGCTAGATTACAAAAATATAT..... 243
   |||
2849  leSerAsnngluYsaAsnslYsaValAlaPheHisThrleuHis 2865
   |||
244  ....AAAAATATAAATGAATAGTAAATCACCACAGACA 281
   |||
2866  AsnmetSerHisLysgluMetSerLysTyrAspLeulleglYsaSlY 2882
   |||
282  AATTTTAGAGAGTAT.....AAAAAGAAAC 310
   |||
2882  sTyrLeuGluasnTyrIleasnasnleuileleugluLysLysLys 2898
   |||
311  AAGCATATTATAGCAGATACGTATGTTCAATGGAACAGTCAAGTGATGA 360
   |||
2899  ....leasnasnleuAsnValHis 2905
   |||
361  GCAAAACATATGATGATTCAAACCAATGTCATATTTATATTCG 410
   |||
2906  IleasnLysLysMetAspAsn.....IleleuYrSerPh 2918
   |||
411  ATTTGCTGCATTAATATGTTATGATGATCATACCATTTATATGAA 460
   |||
2918  eIleasnarGIlleasngluThrArgAspAsnThrLysLysLysL 2935
   |||
461  GA.....CAAGATTCCTATTAAGAAACACGCTTAGAATATGAAC 504
   |||
2935  euTyrIleArgArgTyrTyrleuLysLys...SerIleLysTyrAsn 2950
   |||
505  ....TATGATGATGCTTAATGAATAGACGTACGAAAAAGTACG 548
   |||
2951  HisleuYrAsnMet..... 2955
   |||
549  TCCAGATGTAATATTTTATCTTAGAGTGAGAAACCTATATGATAG 598
   |||
2956  ....ProIlePheleuSerleuPheleuArgCysValThrIleHis 2971

```

```

599  ATATAGAAAAAAGAAATTCATATGTGATTAATAATATGAACTTTAT 648
   |||
2971  IsTyrPheLysPheTyrAsnSerTyrIleTyrPheLysHisTyrAsn 2987
   |||
649  GATTTGATTTAGATGGGTGATATGCTGACGGAACCCAT..... 690
   |||
2988  MetleuHisIle.....ProHisAlaValle 2996
   |||
691  ....GGCAAGTTTACAACT 706
   |||
2996  ulYHisleuYrSerThrHisPheAsnIleasleuPheValasnm 3013
   |||
707  TAAATGAATTA.....AATTTTCAATATATATAT 738
   |||
3013  etleuGluLeuPheTyrValPheIleGluIleTyrAsnAsnTyrPheVal 3029
   |||
739  AAA..... 741
   |||
3030  SerPheCysAspIleSerSerCysArgAsnLysHisValGlnArgAspGl 3046
   |||
742  ....TTAATTAAGCTTTGTAAGAAACCTATTCGGAAGAAAGTTAA 784
   |||
3046  nArgCysleuAsnAsnslYsaAsnLysSerGluAspAsnLysIleT 3063
   |||
785  TT..... 786
   |||
3063  YrCysThrAsnAsnAsnngluYaspclYrAspAspAspGlYrGlu 3079
   |||
787  ....TCATTTCTGCT.....TCATCAAAAGCTGCATATATCATGCT 824
   |||
3080  LysAsnValSerGluIleTyrLysGluAsnAsnslYsaIleAsnVally 3096
   |||
825  TTCAGAGATTCATCTTCTGTAAGATGAAGAAATTCATATTAACATA 874
   |||
3096  sGluAspIleTyrAsnIleasnasnIleasnValTyrProleuAsnnglu 3113
   |||
875  AATTTTGTCTGAAACAATAGAAACAATTAAGAAATTAATGAGGACACA 924
   |||
3113  YsleuValSerIleTyrleuAsnThrleuYsaGluIleLysGluCys 3129
   |||
925  GCGATGTTATCAGCAGCACTTTATTAATTTTATATACGCAAG.. 972
   |||
3130  TyrGluCysHisleuGluHisMetLysAsnAsnslYsaMetGluYsa 3146
   |||
973  ....GAGAAATAGACTTGTATTTATTCAAACATACATTTAG 1012
   |||
3146  rPhePheIleGluHisleuLeuTyrPheLeu.....TyrAsnArgI 3161
   |||
1013  AAAGTACAAATCCAGATATATGATGATATGATATGATATCCATTTAT 1062
   |||
3161  leasnThrIleTyrGluLeuPheTyrAsnPheTyrPheThrTyrLeuArg 3177
   |||
1063  TTTGTTTAAATATTAACATCACATCATATATAGTTTTCATAGACA 1112
   |||
3178  ...LysLysGluAsnAsnAsnAspIleleuLeuAspIleValAsnGluH 3193
   |||
1113  TTAACAGAGGTGATTTAGTCCGAAAAATAAGAAATTAATTAAGATTT 1162
   |||
3193  s.....IleTyrAsnleuIleG 3199
   |||
1163  GAAAAACATATCATGATATA...AATCAAAATATATATAGGCGAGATGT 1209
   |||
3199  LysnLysIleTyrAspGluIleasnlYsaIleasnasnPheleuAsp 3215
   |||
1210  ATAGGATATGCGCAATTTATTTATGAAGAACAATTAACCACTGATCAT 1259
   |||
3216  LysGluTyrTyrTyrPheTyrIleAsn.....ThrleuThrPh 3228
   |||
1260  CGATGTAGATATTTTCTTCAAAATATTTGGAACAATTAATTCCTGAAG 1309
   |||
3228  e..... 3228
   |||
1310  TACAAACTCCAAAGACCTTACTATTAAGTGAACCTGAAAGACTGAGC 1359

```

```

3229 .....| | | | | | | | : : : | | |
1360 ACATAGATGAAATATGTCCAGACCTCGTTATTCCAACCATAGGATTA 1409
3237 .....| | | | | | | | : : : | | |
1410 TTCAACAACATAATGATGCTATATGGAAAACATAGATCTTATTCATTATC 1459
3238 rlellellylys..LysIleleuaasnllyserlleylelrillepropheile 3254
1460 CACCTGGTGTAGACAGATATGATGGGACTGG...TCAAGATGCTAT 1506
3255 TyrHisSerLeuPheaspIySaSnLySasnPherThSerIeYrHisIl 3271
1507 G.....AAAAATATGCATGGGAAGCACGCCATTATT.....A 1541
3271 eaaNaasnaSGlnTyrlleargsnsllysaSpHIsIellePhecysSerl 3288
1542 TAACTACTGACTATAAGAAGAACTCTATTTATATATGGAAGGGGAACAT 1591
3288 eutleValpheIlelleYlelleYtValpheVallys..... 3301
1592 ATTAAATTAAATGGTGGAACAAGAACCTCCGGAAGTCAGCAGCTACAG 1641
3302 .....| | | | | | | | : : : | | | : :
1642 TCATACACAAMACTAGATGCATCCAAATGTCCAGGATAGAAATATGAA 1691
3312 nHISglInASnASnLySaSnHISglInASnVal..... 3322
1692 TAAAAAATATCCACATTAACCAC 1714
3323 ....AsnthraSnValIGlyHIS 3328

seq_name: /SID52/gcgsdata/geneseq/geneseqp/AA2000.DAT: AAB18272
seq_documentation_block:
ID AAB18272 standard; Protein: 2500 AA.
XX
AC AAB18272;
XX
DT 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related protein SEO ID NO:129.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-O107131.
XX
PA (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S., Carucci D., Gardner M., Venter JC;
XX
WP: 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
DS Disclosure; Page 302-309; 577pp; English.
```

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*. Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the plasmodium chromosome 2 and the subsequent identification of the proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic life cycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAT0078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

```
alignment_scores:
    Quality: 136.50      Length: 4855
    Ratio: 0.588          Gaps: 27
Percent Similarity: 47.835 Percent Identity: 21.649
```

```
alignment_block:
US-09-579-383-2 x AAB18272 .
```

Align seg 1/1 to: AAB18272 from: 1 to: 2500

58 TCAGAACCTTGAAAGCAAAAAAATAATATTAATTAATTCATTGGCAATAT 107  
 118 SerAsnAsnLeuValLysAsnGluIleThrAspLys.....GluI 1332  
 1318 SerAsnAsnLeuValLysAsnGluIleThrAspLys.....GluI 1332  
 108 ACGGCAAAATTAATAATTA.....ACGCATCAACGGAAATACATGCT 151  
 1332 eSerGluAsnLysAsnGluLysIleIleGluAsnGluLysIleIleGlu 1349  
 152 CTCTTTTCACATCTGTTAAATTCGAATATAGTAATTTGTGAATATGACAT 201  
 1349 snGluLysValValLys.....AsnGluAsnMetValLys..... 1360  
 202 TATTGGCGAGATGGGTGTACTGTACGTACAAATTACAAAAATAATAAATAT 251  
 1361 .....AsnGluLysValValLys...AsnLysAsnVal 1370  
 252 AATAATAAATGAT.....AGAAATATCCACCAAGACAATTTTAAGCG 292  
 1370 ValLysAsnGluAsnValValGluLysAsnSerThrGhrIleLysLysG 1387  
 293 AGTATAAAAAAGCAAAACAGCTATTATAAGCAGATCTATGATGTCATCG 342  
 1387 LnhIAsnLLeSerMetLeuAsnValProAsnTrpTrpGluAsnAsnThr 1403  
 343 AACAGTCAAGTGTATAGAGCAAAACATATGATGTAATTCAAACCAATGCT 392  
 1404 ArgGluLysAspIleThrAsnAsnAsnAsnLLeSerGluAspProLeuVal 1420  
 393 G.....TCAATTTATATATTCGATTGGTCGCGATTAAATATG 430  
 1420 LAsnGluLLeSerThrLeuSerTrpLysProLysSerThrTrpAsnGly 1437  
 431 TATTATGATGTACTAGACCATTTATATGGAACAACAAGATTCCTATTAGA 480  
 1437 LeSerAsnLleIleGluGluGluLysGluLysLysPheValAsnLys 1453  
 481 AAACACGGCTTGATATGAACAACCATATGGTATGATGCTTAATGAATTAAG 530

```

1454 LysLysAsn.....SerLeuAsnHisIleAs 1462
531 ACCTATTCAGAAAAGTA.....CGTCAGATGTAATT 564
1462 nArgAsnGluLysIleTyrIleGluAspAsnLysAsnTyrIleIleG 1479
565 .....CTTTATCCCTTAGTGAGAA 585
1479 LuAsnIleCysLysCysPheHisPheHisSerLeuGluLeuAsnGly 1495
586 ACCATATGATGATATGAAAGAAATGATATGCGATAAAATTT 635
1496 LeuProGluIleAsnValAsnLys.....AspLys..... 1505
636 GAACCTGTTATGATTTGATTTGATGATGTTA.....G 670
1506 AsnLeuTyrAsnAsnLeuHisIleThrAsnCysLeuLeuPheLysG 1522
671 ATATTCACCTGGAAACACACATGGAGTTTACAACTTAATGAATTAAT 720
1522 LuThrThrGluGluValLeuLysLysPheLeuProAsnAsnGluIleAsn 1538
721 TTTTCATATTTATTTATTAATTAATTAACCTGTTAAGAAAACTATTC 770
1539 IleMetSerTyrTyr.....TyrAsnHisIleLeuTyrArgLeu.A 1552
771 GGAGAAAAGTTAATTTCAATTTCTGGTTCATCAATGCTGCAT...TAT 817
1552 rGmetLysAsnLysTyrGluAspIleIleHisAspLysLeuHisValLys 1568
818 CATGCGTTTCAGAGTTG..... 835
1569 HisLysLeuLysGluLeuIleLysTyrGlnTyrLysGluTyrLeuLeu 1585
836 .....CATCTTCCTGAA 848
1585 sLysThrValTyrProArgAsnIleCysArgAsnGluHisMetAsnGln 1602
849 AG.....ATGAGAAATCCCATATTAACACT..... 873
1602 ysaAspAsnCysThrLysAspIleTyrIleAsnGluAspAsnAsnLysThr 1618
874 .....AAATTTTGTCTGAACAAATAGAAACAATAAAGA 908
1619 GluLeuAsnIleGluLysIleSerLysGluAsnAsnGluAsnLysAs 1635
909 ATTACATAGGCGACAGCATGTATCA.....GCAGGAACCTTTTATA 952
1635 nThrTyrMetAsnThrThrSerTyrLysGluLeuGluLysAsnTyrIleA 1652
953 ATATTTTAAATACA.....GCAAAGAGAGAA 978
1652 snPheLeuAspThrPheAsnLeuTyrAspAsnIleTyrSerLysGluLys 1668
979 ATA.....GATCTGTGA..... 990
1669 TyrGluThrAspGluAsnAspLeuIleLeuAsnAsnLysGluProSerI 1685
991 .....TTTATTCAAACATACAT.....TTAGAAACTACAA 1021
1685 eSerTyrAsnPheAsnSerAsnTyrAsnAsnAspLeuLeuLysSerAspA 1702
1022 AT.....CCAGTATTAATGTAGATATGTAATACTTA... 1050
1702 snValTyrGluTyrIleTyrLysAspIleTyrTyrAspSerTyrTyrAsp 1718
1051 .....TCCCATTTATATTTGGTTAAATATATACATCAACAATCATATT 1094
1719 LysAsnThrTyrIleTyrLysAspAsnLysTyrThrPhe..... 1731
1095 AGGTTTTCATTAGACATAACAGAGGTGATTTCCCGAAATATAG 1144

```

```

1732 .....HisLysThrAsnSerPheIleAsnAspGluAsnG 1743
1145 AATTATTAGAAATGCTAGAGAAAACAATACATGATTAATAATCAAAATPAT 1194
1743 LysCysTyrHisLeuLeuThrTyrProLeuGluAspGluIleGluAsnMet 1759
1195 AAT 1197
1760 Asn 1760
seq_name: /STD2/gcgdata/geneseq/geneseqp/AA1996.DAT:AAW02159
seq_documentation_block:
ID AAW02159 standard; Protein; 866 AA.
XX
AC AAW02159;
XX
DT 14-JAN-1997 (first entry)
XX
DE Soluble chitinase.
XX
KW periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW catabolism.
XX
OS Vibrio furnissii.
XX
PN W09625424-A1.
XX
PD 22-AUG-1996.
XX
PF 13-FEB-1996; 96WO-US02332.
XX
PR 13-FEB-1995; 95US-0386727.
XX
PA (UYJO ) UNITV JOHNS HOPKINS.
XX
PI Bassler B, Chitlaru E, Keyhani N, Roseman S, Rowe C;
PI Yu C;
XX
DR WPI: 1996-39335/39.
DR N-PSDB: AAT36390.
XX
PT Chitin biosynthetic enzymes end I, exo I and exo II - are
PT periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT aryl beta-N-acetyl:gluco:amidase(s), respectively
XX
PS Example 4; Page 79-82; 101pp: English.
XX
CC Periplasmic chitodextrinase (AAW02156), periplasmic
CC Beta-N-acetylglucosaminidase (AAW02157) and aryl
CC Beta-N-acetylglucosaminidase (AAW02158) can be used to produce chitin
CC oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC higher, by contacting them with soluble chitin. The enzymes are
CC encoded by the genes endI, exoI and exoII respectively. They are
CC all genes involved in the catabolic pathway of chitin.
XX
SQ Sequence 866 AA;

```

alignment\_scores:

Quality:	130.00	Length:	359
Ratio:	0.812	Gaps:	13
Percent Similarity:	44.568	Percent Identity:	18.942

alignment\_block:

```

US-09-579-383-2 x AAW02159 ..
Align seq 1/1 to: AAW02159 from: 1 to: 866
211 GATGCGTGT..... 219
118 AspGlyCysThrLeuSerAspLysLysGluIleValValAlaAspThrAs 134

```

[illegible]

CC may help to reduce or eliminate these problems.  
 XX Sequence 866 AA;

alignment\_scores:  
 Quality: 130.00 Length: 359  
 Ratio: 0.812 Gaps: 13  
 Percent Similarity: 44.568 Percent Identity: 18.942

alignment\_block:  
 US-09-579-383-2 x AAY52307 ..

Align seg 1/1 to: AAY52307 from: 1 to: 866

```

211 GATGGGTGT..... 219
|||||
118 ASPGLYCYSYTHLEUSERSPLYSGLIILEVALVALAASPETHRAS 134
..AAGTCAGATTACAAAAATAATAATAATAATAATAATAAGATAGAA 268
|||||
134 POLYSETHISLEUALAPROLEUASNALPROLEUNGINGIUSASNLYS. 150
269 AATCACCAAGACAATTTAGAGAGTATAAAAAAGCAACAGGTAT 318
|||||
151 .....ProTYrTHrAsnLysAlaGlyLysVal 159
319 ATAGCAGATTACTATGTTTCATGAGAACAGTCAGATAGACAAACA 368
|||||
160 VALGIALATYTYRVALGILUTRPGLYVALTYGLY....ArgLysph 174
369 TATGATTGATTCACAAACCAAGTGTCAATTTATATATTCATTGCT. 417
|||||
174 ETHRVALASPLYSILEPROALALYASNLLEUTHRHISILEUTYGLYP 191
418 .....CGCATTAAT..ATGTTATATGATGA 441
191 heThPrOleCysGLYGLYAsnGLYILEAsnSPserLeuLysGLUile 207
442 TCTAGACATTTAATGAGAACAAAGATTCCTATTAGAAACAGCGCT 491
|||||
208 SERGLYSERPHGLUALALEUCLINARGSERCYSIAGIARGGLIASPH 224
492 ACAATAT..... 498
224 elYsValSerIleHISAspPrOTRPaLaLaValGImetGLYInGLY 241
499 .....GAACCTATGATGATGATGCTTAATGAATTTAGA 531
241 snLeuThrAlaPheAspGIuPrOTYrLysGLYAsnPhelGLYAsnLeuMet 257
532 CGTATCAGAAAAAGTACGTCACAGATGTAATTTATCTTTATCCTTAGTGG 581
|||||
258 ALALEULYSYLAALAsnPROAsnLeuLysIleLeuProSerValGLY 274
582 AGAAAC.....TATATGATAGATAGAAAAAGAA 613
|||||
274 YTRPhrLeuSerAspPrOphETyrPhePheSerAspLysThrLysArg 291
614 TTGATTATGCTGATTAATATTTGAAGCTTTGAATGATTTGAT...TTA 660
|||||
291 sPTThrPheValAlaSerMetLysGLUTYrLeuGLInThrTrpLysPhe 307
661 GATGGTGTAGATTTGACTGGAA.....CCACATGGCAAGTTTGA 701
|||||
308 ASPGLYVALASPILEAspTRpLuphePrOGlyGLYInGLYAlaAsnPr 324
702 CAACCTAAATGAATTAATTTTCAATTAATTAATTAATTAATTAATTA 751
|||||
324 AsnLeuGLYGLYPrOAsnAspGLYAlaThrTYrValAlaLeuMetLysG 341
752 TGTTAGAAAAAATTCGGAAGAAAGTAAATTCAATTCTGTGTCA 801
|||||

```

```

341 IuLeuArgAlaMetLeuAspGIu..... 348
802 TCAAAATGCTGATTTATCATGCGTTTCAGAGTTGCATCTTCTGTAAAGA 851
348 ..... 348
852 TGAAGATCTCCATATACACTAAATTTTGTCTGACAAATAGAAACA 901
349 .....LeuGluAlaGIuThG 354
902 ATAAAGATTACATFAGGCGAGCGATGTTATTCAGCAGCACTTTAT 951
|||||
354 IYArgInTYrGIuLeuThrSerAlaIleSerAlaGLYAspLysIle 370
952 AAT.....ATTTTAATACAGCAAGAGAAATAGATCTGTATTAT 995
|||||
371 ALAlYSValAspTYrGLuAlaAlaGLInGLYrMetAspTYrIlePhe 387
996 TCAAAACATTCAT.....TTAGAAACTACAAATCCAGATATRA 1033
|||||
387 uMetSerTYrAspPheSerGLYAlaPheAspLeuLysAsnLeuAlaHISG 404
1034 TGTGATATGTACTTATGCCATTATATTGTTTAAATAT..... 1077
|||||
404 InThrAsnLeuTYrAlaSerSerTrpAspPrOAlaThrLysTYrThr 420
1078 .....AACATCAACATCATATTAGGT 1098
421 AspLysGLYValLysAlaLeuLeuGLY 429
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA52533
seq_documentation_block:
ID AAY52533 standard; Protein: 509 AA.
XX
AC AAY52533;
XX
DT 22-FEB-2000 (first entry)
XX
DE D. pteronyssius 98 kD mite allergen protein (map) Pderp98-509.
XX
KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;
house dust mite; IgE; immunoglobulin E; allergen; mapB;
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
canine; veterinary; antibody; vaccine; immunisation.
XX
OS Dermatophagoides pteronyssius.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /note- "Signal peptide"
FT Protein 20..509
FT /note- "Mature Pderp98-509"
XX
PN W09954349-A2.
XX
PD 28-OCT-1999.
XX
PF 16-APR-1999; 99MO-US08524.
XX
PR 17-APR-1998; 98US-0062013.
PR 13-MAY-1998; 98US-0085295.
PR 02-SEP-1998; 98US-0098909.
XX
PA (HESK-) HESKA CORP.
XX
PI McCall CA, Hunter SW, Weber ER.
XX
WP1; 2000-052700/04.
DR N-PSDB; AA238585, AA238586, AA238587, AA238588.
XX
PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides
used to modify an animals' hypersensitivity to mite allergens

```

XX Claim 3: Page 134-136; 154pp; English.

PS This sequence represents Dermatophagoides pteronyssinus mite allergen  
XX protein (map) pDerp98-509. pDerp98-509 has a molecular weight of 98 kD,  
CC comprising 509 amino acids, and has a high degree of homology with the  
CC D. fariniae 98 kD allergen, mapB (AAV52523). Nucleic acid molecules  
CC encoding pDerp98-509 were isolated from a D. pteronyssinus cDNA library  
CC by hybridisation with a probe encoding the D. fariniae high molecular  
CC weight map (HWM-map) composition. Mite allergenic proteins and  
CC peptides, and nucleic acids encoding them, may be used in therapeutic  
CC compositions to modify an animal's hypersensitivity reaction to mite  
CC allergens. Animals that may be treated include mammals and birds,  
CC especially felines, canines, equines, humans, other pets, and work or  
CC diagnose allergies via a skin test. The proteins and peptides can also  
CC be used to raise antibodies, which have a variety of potential  
CC uses. For example, they can be used as vaccines to passively immunise  
CC animals against dust mite hypersensitivity, as positive controls in  
CC test kits and as tools to recover desired dust mite allergens from a  
XX mixture of proteins.

Sequence 509 AA:

alignment\_scores:                   Quality: 126.50                   Length: 519  
                                      Ratio: 0.583                   Gaps: 25  
Percent Similarity: 41.811           Percent Identity: 18.304

alignment\_block:  
US-09-579-383-2 x AAV52533   ..

Align seg 1/1 to: AAV52533 from: 1 to: 509

```
193 TATGATCTTTTGGGAGATGGTGCTACTAGATTAATTAACAAAATTA 242
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5 PheAlaLeuPheCysIleTrpAlaCysIleGlyLeuMetAsnAlaAla 21
243 TAAATAATTAATAATAAATGATAGAAATCACACAGACAATTTTAGAG 292
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
21 rLyAsrArgAspHisAsnAsnTYrSerLyAsnProMetArgIleVal... 36
293 AGTATAAAAAAGAAACAGAGTATTATAGACAGATACTATGCTTCATGG 342
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
37 .....CysTYrValGlyThrTrp 42
343 AACAGTCAGGTGATAGACAAACATATGATTGTTCAACCCCATGGT 392
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
43 Ser.....ValTYrHisLySValAspProTYrThrIleG1 54
393 GTCAAAT.....TTATATATGTCATTTGCTGCGCA 421
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
54 uAspIleAspProPheLySCTyrHisIleuMetCylrPheAlaIysI 71
422 TTAAT.....ATGTTATATGATGTATCTAGACCATTTAATGAAGA 462
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
71 leAspGIuTYrLySCTyrThrIleGlnValrPheAspProPheGlnAsp 87
463 CAAAGATTCCTATTAGAAAACAGGCTTAGAATATGAACCTATGTAT 512
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
88 AsnHisAsnSerTrpGIuLySHisGIyTYrGIu..... 98
513 GATGCTTAATTAAGATTAAGCTATACGAAAGTAGTCACAGTGAATTA 562
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
99 .....AlrPheAsnAsnLeuAlrGleuLySAsnProGIuLeuThrT 112
563 TTCTTTATCTTAGTGTAGAAAACCTATATGATATAGATAAAAGAA 612
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
112 hrMetIleSerLeuGIyGIy.....TrpTYrGIuGIySerGIuLySTrp 126
613 ATTGAT.....TATGTGATAAAT 632
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
```

```
127 SetrAspMetAlaAlaAsnProThrTYrArgGlnGlnPheValGlnSerVa 143
633 ATTGAAGCTTGTTAATGATTTTGATTTAGATGGGTAGATTAAGCTGG 682
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
143 lLeuAspPheLeuGlnGIuTYrLySAspPheAspLyLeuAspIleuSpr 160
683 AACACATGAGGAAGTTTACAACTTAATGAATTAATTTTCAAAATAT 732
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
160 lUTyrProGIySerArgLeuGIyAsnProLySleAspLyGIuAsnTYr 176
733 TATATTAATTAATTAATTAATTTTAAAGAAAACCTATCCGAGAAAAGT 782
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
177 .....LeuThrLeuValAlrGIu..... 182
783 AATTTCAATTTTGGTTCATCAAAATGCTGCATATATCATGCTTCAGAG 832
182 ..... 182
833 TTGCATCTTTCTGTAAGATGAGAAATCTCATATACACTAAATTTTGG 882
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
183 .....LeuLySGIuAlrPheGIuProPheGIy..... 191
883 TCTGAACAATAGAAACAATAAATAATTAACATAGGACGACGATGTT 932
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
192 .....TYrLeuLeuThrAlaAlaVa 198
933 ATCAGCAGAACTTTTATTAATATTTTAAATACAGACAGAGAAATAG 982
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
198 lSerProGIy.....LySAspLySleA 206
983 ATCTGTATTT..... 993
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
206 spValAlaIyTrGIuLeuLySGIuLeuAsnGlnLeuPheAspTrpMetAsn 222
994 ATTCAACATCAATTTAGAA..... 1014
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
223 ValMetThrTYrAspTYrHisGIyGIyTrpGIuAsnValrPheGIyHisAs 239
1015 .....ACTACAATCCAGATATATAGTGTATGATGACTTATATCCC 1054
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
239 nAlaProLeuTYrLySAsrProAspGIuThrAspGIuLeu.....H 253
1055 ATTATATTTTGGTTTAAATATTAACATCAATCAATATTAAGTTTTC 1104
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
253 lSthTYrPheAsnValAsnTYrThrMetHisTYrTYrLeu..... 266
1105 TTAGAACATACAGAGGTGATTAAGTTCGCCGAAATATAAGATATTA 1154
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
267 .....AsnAsnGIyAlaThrArgAspLySLeuValrMetGIyValr 280
1155 ATTGGTAGAATAA.....ACAAATACATGATTAATAATCAAAATTAATA 1198
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
280 oPheTYrGIyAlrGAlaTrpSerIleGIuAspArgSerLySValLySLeu 296
1199 GGGCAGATGTATAGGATATGCA.....TTTATTTAT...GA 1236
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
297 GIyAspProAlaLySGLyMetSerProProGIyPheIleThrGIyGIu 313
1237 GAACATTAACCAACTGATGATTCATGATGTAGATATTTTCTTAACAATAT 1286
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
313 lUGlyValLeuSerTYrThrIleGIuLeuCysGlnLeuPheGlnLySGLu 339
1287 TTGGAACATTTTAATCTCGAAGTAAACAACATCCAAAGACCTTACTATA 1336
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
329 uTrp...HisIleGln..... 333
1337 CTGAACACCTGAAGACTGTAGACAAATAGATGATATGTGCCAGACTC 1386
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
334 .....TyraSpGIuTYr..... 337
1387 GTATATCCAACATAGGATATATTACAACACATAGATGCTATATGAA 1436
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
338 .....TYrAsnAlaProTYrGIyTYrAsnAspLySleTrp.. 349
```

Tue Mar 26 09:19:07 2002

```
1437 AACTAGATCTTATTCATTGACCTGGGTAGACAGATATGATGGG 1486
349 ..... 349
1487 ACTGGTCAAGTATGCTATGAAAAATATGCGATGGGAAGCAGCCCAT 1536
      ||| |||::: ::::: :::::
350 .....ValGlyTyrAspAspLeuAlaSerIleSerCysLysLeu 362
1537 TATTATACTGACTATTAAGAAAGCTCTATTATTATATGGAAGGGA 1586
      ::| ::| ::|::: ::|::: ||
363 AlaPheLeuLysGluLeuGlyValSerCysValMetIleTyrPheLeuGlu 379
1587 A 1587
379 u 379
```









```

;
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-566-347-2

```

```

alignment_scores:
  Quality: 142.50      Length: 425
  Ratio: 0.720        Gaps: 20
  Percent Similarity: 46.588      Percent Identity: 20.471

```

```

alignment_block:
US-09-579-383-2 x US-08-566-347-2 ..

```

```

Align seg 1/1 to: US-08-566-347-2 from: 1 to: 561

```

```

25 TTAATTAATGATCCATCTGTATTCGCAAAATTCAGAACCTGGAAGG 74
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
11 LeuLeuIleGlySerThrLeuCysSerAlaAlaGlnAlaAlaAlaProGln 27
75 AAAAAATTAATATA.....AATAATTCATGGGAATAATACGGGAAA 115
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
27 YLysProThrIleAlaTrpGlnAsnThrLysPheAlaIleValGluValA 44
116 AATAAATAAATCAATCAACGAAATACATGAGCTTTTTCACATCTT 165
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
44 spGlnAlaAlaThr.....AlaTrpAsnAsnLeuValLysVal 56
166 AATCGAATATATTAATTTGTAGAAATATGATCTTATTCGGAGAT.. 213
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
57 LysAsnAlaAlaAspValSerValSerTrpAsnLeuTrpAsnGlnLysPAl 73
214 GGGTGTAACTCTAGAATTACAAAAAATAATATA..... 246
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
73 agLYThrGlyProLysIleLeuLeuAsnGlnLYsGluAlaTrpSerGlyP 90
247 .....AATATAAATAAATGAT 264
90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLYsGlyGly 106
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
265 AGA..... 267
107 ArgGlyTrpGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123
267 ..... 267
123 rAspAlaThrGlnIleValValAlaAspThrAspGlySerHisLeuProp 140
268 .....AATCCCAAGACAAATTTTAGAGGAGTATAAAAAAGAAACAA 312
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
140 rIleuLYsGlnPro.....LeuLeuGlnLYsAsnLYsProTrpLYsGln 154
313 GGT.....ATTATAGCAGATACATGATGTCATGAGAACACATCAAG 353
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
155 AsnSerGlyLysValValGlySerTrpPheValIuTrpGlyValLYrGln 171
354 TGATAGACAAACATATGATGATCAACCAATGCTGCTCAATTTAT 403
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
171 Y.....ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThr 186
404 ATATTCGATTTGCTCGCATT.....AATATGTTATATGAT 438
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
186 IsteuLeuTYrGlyPheIleProIleCysGlyGlnGlyIleAsnAsp 202
439 .....GTATCTAGACCATTTAA 455
203 SerLeuLYsGlnIleGlnGlySerPheGlnAlaLeuGlnAspSerCysGln 219
456 TGAAGACAAAGATTCCTATTAAAGAAAAC.....GGCTTAG 493
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
219 nGlyArgGlnAspPheLysIleSerIleHisAspProPheAlaIleLeuG 236
494 AATATGAACCTATGTAATGATG..... 516

```

```

236 InLYsAlaGlnLYsGlyValThrAlaTrpAspAspProTrpLYsGlyAsn 252
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
517 CTTAATGAATTTAGACGTCATCGAAAGTAGTCACGATGTAATTTCT 566
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
253 PheGlyGlnLeuMetAlaLeuLYsGlnAlaHisProAspLeuLYsIleLe 269
567 TTTATCTTACGTGAGAAAC.....TATATGATG 598
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
269 uProSerIleGlyTrpThrLeuSerAspProPhePheMetGlyA 286
599 ATATAGAAAAAGAAATGATTTGTGGTAAATATATGACCTGTATAT 648
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
286 sPLysValLYsArgAspArgPheValGlySerValLYsGlnPheLeuGln 302
649 GATTTTGTAT...TTAGATGCTGTAGATATGATGCTGGAAA...CCACATG 692
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGlnPheProGlyGln 319
693 GAAGTTTTC.....AAGTTAAATGAATTAATTTTCAATATATATA 736
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
737 TTAATTAATTAATCTGTATAGAAAAACTATCCGGAAGAAAGTAAAT 786
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
336 aLeuLeuMetLYsGlnLeuArgAlaMetLeuAspGln..... 348
787 TCAATTTCTGCTCATCAATATGTCGATTAATCATGCGTTTCAGAGATGC 836
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
349 .....LeuSerAlaGlnThrGly..... 354
837 ATCTTTCTGTAAAGATGAAGAAATCTCCATATTAACATAAATTT..TTGT 883
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
355 .....ArgLYsTrpGlnLeuTr 360
884 CTGAACAATTAAGAAACAATTAAGATTAATACATAGGCGACGCGCATGTA 933
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374
934 TCACACAGAACTTTATTAATATTTTATATACAGCAAGAGAAATATGA 983
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
375 .....TyrAsnValAlaGlnAsnSerMetAs 383
984 TCTTGATTTATTCAAACATACAT 1008
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
383 rHisIlePheLeuMetSerTYrAsp 391

```

seq\_name: /cgn2\_6/prodate/2/1aa/5A\_COMB.pep:US-08-693-835-2

seq\_documentation\_block:

Sequence 2, Application US/08693835

Patent No. 5776448

GENERAL INFORMATION:

APPLICANT: Suslow, Trevor V.

APPLICANT: Jones, Jonathan D.G.

TITLE OF INVENTION: No. 5776448el Chitinase-Producing Plants

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESSES:

ADDRESSER: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/693,835

FILING DATE: 01-AUG-1996

CLASSIFICATION: 435



```

; Patent No. 5792647
; GENERAL INFORMATION:
; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASLER, BONNIE
; APPLICANT: KEYMANT, NEMAT O.
; APPLICANT: CHITLARI, EDITH
; APPLICANT: ROME, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/386,727
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOBBS, ANN S.
; REGISTRATION NUMBER: 36,830
; REFERENCE/DOCKET NUMBER: 4130/206916
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ. ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-386-727-8

alignment_scores:
    Quality: 130.00      Length: 359
    Ratio: 0.812        Gaps: 13
    Percent Similarity: 44.568    Percent Identity: 18.942

alignment_block:
US-09-579-383-2 x US-08-386-727-8 ..
Align seg 1/1 to: US-08-386-727-8 from: 1 to: 866

211 GATGGGTCT..... 219
|||||
118 Aspglycylserleuserasplysluilevalalaaaspthras 134
220 AACCTGAGATTACAAAAAATAATATAATAATAAATGATGAA 268
:::||||:||||:||||:||||:||||:
134 pelyserleuallaproleuasnalaproleuglnlunasnls 150
269 AATCACAACAATAATTTAGAGAGATATAAAAAAGCAAGATTT 318
||| :::::
151 .....Protyrthrlnslsalaaglyllysval 159
319 ATAGCAGATCTATGTTTCATGACAGTCAAGTGTATAGACAAACA 368
:::||||| :::|||||
160 Valglyalatyrtyvalglutpdlvaltyrgly.....Arglysph 174
369 TATGATTATTCAAACCAATGTCATATTTATATATGCAATTTGCT 417
:::||||| :::|||||
174 ethvalasplyslleproalalysasnleuthisileuetyrglyp 191

```

```

418 .....CCATTAAAT..ATGTTATATGATGTA 441
191 hethrproillecysglyslnglylileasnaspsersleuylsluile 207
442 TCTACACCATTAATGGAACAAAGATTCCATTATTAAGAAACAGGGCT 491
||| |||:||||:|||||
208 Serglyserphleglualaleuclnarysercysalaglyarglualsph 224
492 AGAATAT..... 498
224 elysvalserllehisaspprotpralalavalclnmetglynglyla 241
499 .....GAACCTATGGTATGAGCTTAATGAATTAAGA 531
241 snleuthralapheaspluprottyrlyslgysnphleglysnleumet 257
532 CGTATCAGAAAAAGTCCAGATGATATATCTTTATCTTGTGGG 581
:::||||:||||:||||:|||||
258 Alaleuylslsalaasnproasnleuylsileuoproservalglycl 274
582 AGAAACC.....TATATGATGATATTAAGAAAGAA 613
| |||
274 YTPthreuserasppropheTyrrpheheseraplysthrlysaTga 291
614 TTGATTATGTGATTAATAATTTGAAGCTGTTAAATGATTTGAT...TGA 660
:::||||:||||:||||:||||:
291 spthrphelvalalasermetlysluTyrrleuglnthrtprysphe 307
661 GATGCTGATATTTGACTGGGA.....CCACATGGGAATTTTA 701
|||||
308 Aspglyvalasplleasprpdlneproglglynglylalaasndr 324
702 CAACCTTAATGATTAATAATTTTCAATTAATTAATTAATTAATTAAT 751
|||||:||||:||||:||||:
324 casnleuglyglyproasnaspglyalathrttyrvalalaleumetlysg 341
752 TGTTAAGAAAACTATTCGGAAGAAAGTTAATTTCAATTTCTGTTCA 801
|||||
341 Iuleuarglalmethleuasplu..... 348
802 TCAATGCTGCATTATTCATGCGTTTCAGAGATTCATCTTTCTGTAAAG 851
348 ..... 348
852 TGAAGATCTCCATTAACACTAAATTTTGTGACAAATAGAAACA 901
349 .....leuglualagluuthrg 354
902 ATAAAGATTACATFAGGCGAGCGAGATGTTATCAGACGAATTTTAT 951
:::||||:||||:||||:|||||
354 lyarglnlyrglueuthrseralaleaseralaglylaspyslle 370
952 AAT.....ATTTTAATACAGCAAGAAAGAAATAGATCTTGATTTAT 995
:::||||:||||:||||:|||||
371 Alalysvalaspyrglnalalaglnlntyrmetaspyrrllephele 387
996 TCAAAACATACAT.....TTAGAAACTCAATCCAGATATNA 1033
:::||||:||||:||||:|||||
387 umeserlyraspheserglyalalpheaspsleuylsasneulahlsg 404
1034 TGGTAGATATGTACTTATCCATTTATATTTGTTTAAATAT..... 1077
:::||||:||||:||||:|||||
404 luthnasnleuthralaserseTrrpasproalathrlsyrtyrthr 420
1078 .....AACATCACATCATATAGGT 1098
421 Asplysglyvalylsalaaleuclly 429
seq_name: /cgn2_6/ploddata/2/1aa/5h_COMB.pep:US-08-600-452A-8
seq_documentation_block:
; Sequence 8, Application US/08600452A

```

```

; Patent No. 5985644
; GENERAL INFORMATION:
; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASLER, BONNIE
; APPLICANT: KEYHANT, NEMAT O.
; APPLICANT: CHITLAPU, EDITH
; APPLICANT: ROME, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,452A
; FILING DATE: 13-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07662/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-600-452A-8

alignment_scores:
    quality: 130.00      Length: 359
    Ratio: 0.812         Gaps: 13
    Percent Similarity: 44.568    Percent Identity: 18.942

alignment_block:
US-09-579-383-2 x US-08-600-452A-8 ..
Align seg 1/1 to: US-08-600-452A-8 from: 1 to: 866

211 GATGGGTG..... 219
118 Asp61ySthrLeuSerAspLysLysGluIleValAlaAspThrAs 134
220 .AACTGAGATTACAAAATAATATAATAATAATAATGATGAGA 268
134 pGlySerHisLeuAlaProLeuAsnAlaProLeuGlnGlnAsnLys 150
269 AATCACCAGACAAATTTTAGAGAGATATAAAAAAGGAACAAGTATT 318
151 .....ProTyrThrAsnLysAlaGlyVal 159
319 ATGACGAGACTATGTTTCATGACAGTCGATAGACAAACA 368
160 ValGlyAlaTyrTyrValGluTyrGlyValTyrGly.....ArgLysP 174
369 TATGATTGATTCAACCAACCAATGTCATATTTATATTCATTTGCT 417
174 ethrValAspLysIleProAlaLysAsnLeuThrHisIleLeuTyrGlyP 191

```

```

418 .....CCGATTAA...ATGTTATGATGTA 441
191 heThrProIleCysGlyLysAsnGlyIleAsnAspSerLeuLysIle 207
442 TCTAGACCACTTAAATGAGACAAAGATTCTTAAAGAAACAGCGCTT 491
208 SerGlySerPheGluAlaLeuGlnAsnSerCysAlaGlyArgGluAsp 224
492 AGAATAT..... 498
224 elysValSerIleHisAspProTyrAlaAlaValGlnMetGlyGlnGly 241
499 .....GAACCTATGCTATGATGCTTAAATGAATTGA 531
241 snLeuThrAlaPheAspGluProTyrLysGlyAsnPheGlyAsnLeuMet 257
532 CGTATCAGAAAAGTACGTCACATGATATATCTTTATCTTATAGCTGG 581
258 AlaLeuLysLysAlaAsnProAsnLeuLysIleLeuProSerValGly 274
582 AGAAAC.....TATATGATAGATATAGAAAAGAAA 613
274 YThrPThrLeuSerAspProPheTyrPhePheSerAspLysThrLysArg 291
614 TTGATTATGCTGATAAAATATTGAAGCTTGAATGATTTTGAAT...T 660
291 sPThrPheValAlaSerMetLysGluTyrLeuGlnThrTyrLysPhePhe 307
661 GATGCTGATATGATGACTGGGA.....CCACATGGGAGCTTTA 701
308 AspGlyValAspLysPheAspTyrPheProGlyGlyGlnGlyAlaAsnPr 324
702 CACTTAAATGAATTAATTTTCAATTAATTAATTAATTAATTAATTA 751
324 onsnLeuGlyLysProAsnAspGlyAlaThrTyrValAlaLeuMetLys 341
752 TGTTAAGAAAACCTATTCGCGAAGAAAGTTAATTCATTTCTGCTTCA 801
341 IuLeuArgAlaMetLeuAspGlu..... 348
802 TCAATGCTGCATTTATCATGCGTTTCAGAGTGTGATCTTCTGTAAAGA 851
348 ..... 348
852 TGAAGATCTCCATTAATACACTAAATTTTGTCTGACCAATATAGAACAA 901
349 .....LeuGluAlaGluThrG 354
902 ATTAAGAAATTACATAGGCGACGACGATGTTATCAGCAGAACTTTAAT 951
354 IArgGlnTyrGluLeuThrSerAlaIleSerAlaGlyLysAspLysIle 370
952 AAT.....ATTTTAATNCAGCAAGAGAAATATAGATCTTGATTTAT 995
371 AlaLysValAspTyrGlnAlaAlaGlnGlnTyrMetAspTyrIlePhe 387
996 TCAACATACATCAAT.....TTAGAAACTACAAATTCACAGATATAA 1033
387 uMetSerTyrAspPheSerGlyAlaPheAspLeuLysAsnLeuAlaHisG 404
1034 TGTGATATGTTACTTATCCATTTATATTTTGGTTTAAATAT..... 1077
404 IuThrAsnLeuTyrAlaSerSerTyrProAlaThrLysTyrThrThr 420
1078 .....AACATCACATCATATTAGCT 1098
421 AspLysGlyValLysAlaLeuLeuGly 429

seq_name: /cgn2_6/prodata/2/laa/6A_COMB.pep:US-08-929-329-6
seq_documentation_block:
; Sequence 6, Application US/08929329

```

```

? Patent No. 6120770
? GENERAL INFORMATION:
? APPLICANT: Adams, John H
? APPLICANT: Dalton, John P
? APPLICANT: Kapee, Stefan
? TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
? NUMBER OF SEQUENCES: 23
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Barnes & Thornburg
? STREET: 11 S Meridian
? CITY: Indianapolis
? STATE: Indiana
? COUNTRY: USA
? ZIP: 46204
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/929,329
? FILING DATE:
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Breen, John P
? REGISTRATION NUMBER: 38,833
? REFERENCE/DOCKET NUMBER: 835910-28685
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (317) 231-7745
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 934 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
? MOLECULE TYPE: peptide
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE: N-terminal
? ORIGINAL SOURCE:
? ORGANISM: Plasmodium berghei
? US-08-929-329-6

```

```

alignment_scores:
    Quality: 126.50      Length: 504
    Ratio: 0.591        Gaps: 19
    Percent Similarity: 42.460      Percent Identity: 18.651

```

```

alignment_block:
US-09-579-383-2 x US-08-929-329-6 ..

```

```

Align seg 1/1 to: US-08-929-329-6 from: 1 to: 934

```

```

70 AAAGAAAATAATATAATATTCATTGGGAATA.....ATACG 110
   :::::|||||  |||  |||:::|||||:::
394 GlnAsnLysLeuAsnSerMetAsnSplLeuGlyValCysValLeuLeu 410
111 GGAAATAATAAATACTCATCAACGGAATACATGAGCTTTTTCAC 160
   ::|||:::  :::  :::  :::  |||:::
410 sGluArgProThrCysValLeuLysGlnAsnTyrTyrSerPheTha 427
161 AACTTAATTCGAAT.....AATAGTAATTTGTAGAAATAGAGCT 201
   ::|||  ::|||  |||:::  ::|||  :::
427 snLeuThrAlaAsnTyrPheAspAsnGlnAsnIleGluTyrProAsp 443
202 TAT.....TCGCGAGATGGGTCTAAGCTAGCAATTACAAAAA 239
   |||  :::  |||:::  :::  |||
444 IleGluAsnValLysIleGlyLysAsnArgAsnSerGluLeuSerGlyAs 460
240 T.....ATAATAATAATAATAAATG 262

```

```

460 nLeuThrTyrAsnAspLysLysPheLysAsnSerAsnIleAsnLysGlyM 477
263 ATAGAAAATCACCACCAAGACAAATTTAGAGAGCTATTAAGAAAAAGAAACA 312
   |||  :::::  |||  |||  :::::  |||
477 etValIleSerMetAsnSplIleSerGluIleLysGluAsnSerLys... 492
313 GGTATTATAGCAGATACTATGTCATGACGACAGCAAGTGAATAGAC 362
   :::  :::  :::  :::  :::  :::
493 .....LeuGluThrAsnLysArgAsnGluGlyLysLys 504
363 AAACATATGATGATTCAACCAATGGTGCATTTATATATTCAT 412
   ::|||:::  :::  :::  |||:::  |||  |||:::
504 rLysTyrGlyLeuTyrAsnTyrProIleThrProIleSerTyrLeu... 519
413 TTGCTCGCATTAATATGATATGATGATCTAGACCATTAATGAGACA 462
520 .....GlnIleAsnHisLysArgGluLeuThrLysLysTyrMetTyrPhe 534
463 CAAGATTCCTATTAGAAAACAC.....GGCTAGAAATAGA 500
   ::|||:::  |||  ::|||
535 GluAsnSerLeuThrProSerHisAsnThrGlyAlaSerIleArgTyrGly 551
501 A.....ACCTATGCTATGATGCTTAATGAATTAACATCAGAA 541
   |  |||:::  |||:::  |||:::
551 uGlyAsnAsnArgLeuGlyIleAspAlaAsnAsnAsnArgAsnThrT 568
542 AAGTAGCTCCAGATGATTAATTTCTTTATCCTTGGTGGAGAACCTAT 591
   |||:::  :::  :::  :::  :::
566 yGlyGlyThrGlnAspIleAsnLeuLysArgAsnAsnAsnTyrLysGlnPro 584
592 ATGATAGATATAGAAAAGAAATGATTATGATGATTAATTAAGAGCT 641
   :::::  :::  :::  |||:::  |||:::
585 LysAsnLysLeuAsnProGlnIleGluTyrMetAspArgPhe..... 598
642 TGTATATGATTTTGTATTTAGATGGGTAGATATTTAGCTGGACACAG 691
   |||:::  |||:::  :::  |||:::  |||:::
599 .....AspIleGluLysAsnHisIleTyrIleAspIrrLysGlnAspG 613
692 GGAAG..... 696
613 LysLysTyrGlySerAspLysLeuLysTyrAsnIleIleSerHisGluThr 629
696 ..... 696
630 AlaAsnThrValGlnSerLeuLeuIleThrAspLysAsnAspIleCysPr 646
696 ..... 696
646 AsnHisTyrSerProGlyArgAlaGlnGlySerCysProAsnTyrGlyL 663
697 .....TTTAC 702
663 ySerIleIleValLysAlaLeuGluGlyThrAsnGlyAspGluTyrPhe 679
703 AACTTAATGAATTAAT.....TTTCAATTTATATAT 737
   |||:::  |||:::  |||:::  |||:::
680 AsnLeuAsnPheLeuAsnGluIleArgThrGlyTyrLeuAsnArgTyrMe 696
738 TAAATTA..... 744
696 LysTyrAspValGluLeuProTyrGlyLysSerGlyLeuAlaMetHisH 713
745 .....ATTAAGTTGTTAAGAAAACATATTCGGAAGAAAGTTAAT 786
   ::|||  ::|||  |||:::  |||:::
713 LsGlyAspLeuAsnGlnCysProLysSerLeuAspGlnClnAsnLeuTyr 729
787 TCAATTTCTGCTTCATCAAAATGCTGCATTTATCATGCGTTTCAGAGACTGC 836
   ::|||:::  ::|||  ::|||
730 LysIleLysSerAspLysTyrAsnTyrGlyMet..... 739
837 ATCTTTCGTAAAGATGAAGATCTCCATATACATAATTTTGTGTCG 886
   |||:::  |||:::  |||:::

```





```

241 LysnIleasprhLeuLysTyValCysasnHispheIleArgValIle 257
673 .....ATTGACTGGCAACCATGAGAGTTTTCACATTAA 710
258 LysLeuAsnAsnIleThrTy.....AsnTyrlYsAsnLysGI 270
711 TGAATTAAATTTTCAATTATTAATTAAATTAATTAATTAATTAAGAA 760
270 uGIupheAsnTyrlThrasnGIumetIle.....Thrasna 282
761 AAACATATCCGGAAGAAAGTTAATTCATTTCTGCTCATCAATGCT 810
282 sPserSerMetGIuAspHisAspAsnGIuIleAsnGIySerIleSerAsn 298
811 GCATTATCATCGCTTTCAGAGATGCATCTTCTGTAAGATGAA...GA 857
299 pheAsnAsnCysProser...IleSerSerpheAspLysSerGIuSerLy 314
858 ATCTCATATACACTAAATTTTGTCTGAACAATA..... 894
314 sAsnValIleAsnHisThrLeuLeuArgAspLysMetAsnLeuIleThrs 331
895 .....GAACAATAAAGAAATTACATAGCGACAGCATGTTATCA 936
331 erSerGIuGIuTyrlLeuLysAspLeuHisAsnCysAsn..... 343
937 GCAGAACTTTTATTAATTTTAAATACAGCAAGAGAAATAGATCT 986
344 .....PheserAsnSerSerAspLysAsnAsp.. 352
987 TGTATTTATTCAAACATACATTTTGAAGAACTACAAATCCAGATTAATG 1036
353 .SerpheheLysLeuTyrgIyIle.....CysGIuT 363
1037 TAGATATGACTATATCCATTAA..... 1059
363 yAspLysTyrlLeuIleAspLeuGIuAsnAlaSerPheHisTyrlAsn 379
1060 .....TATTTTGTTTAAATATTAATACATCACAAATCAT 1091
380 AsnValAspGIuTyrgIyTyrlTyrlAspValAsnLysAsnThrAsnIleLe 396
1092 ATTGAGTTTTCATTAGACATTAACAGAGGTGATTAAGTCCCAAAATA 1141
396 uSerAsnAsnLysIleGIuGIuAsnAsnAsn.....GIuAsnA 410
1142 AAGAATTATTGAATTTGTAGGAAAAACAATACATGATTAATAATCAAAAT 1191
410 snLys.....AsnAsnLysAsnAsnAsn 417
1192 AATAATAGGCGAGATGTATATA 1212
418 AsnAsnGIuValAspTyrlIle 424

```

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-150-741-2

seq\_documentation\_block:

Sequence 2, Application US/09150741

Patent No. 6183996

GENERAL INFORMATION:

APPLICANT: Stewart et al.

TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate

Patent No. 6183996

TITLE OF INVENTION: Synthetase II

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/09/150,741

EARLIER APPLICATION NUMBER: PL6380

EARLIER FILING DATE: 1992-12-16

EARLIER APPLICATION NUMBER: A093/00617

EARLIER FILING DATE: 1993-12-02

EARLIER APPLICATION NUMBER: 08/446,855

```

; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRt
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

alignment_scores:
  Quality: 118.00      Length: 457
  Ratio: 0.554         Gaps: 22
  Percent Similarity: 46.608      Percent Identity: 20.569

alignment_block:
US-09-579-383-2 x US-09-150-741-2 ..

Align seg 1/1 to: US-09-150-741-2 from: 1 to: 2391

46 TATCTCGCAATTCAGAACCTTGAAAGAAAAAATAATATTAATTAATTC 95
|||||.....:.. |||||.....:..
42 TyrlSerValGIyTyrgIuGIyCysLysGIyAsnAsnSerIleSer.... 56
96 ATTGGAAATAATACGGGAAATAAATAAATCAATCAACGGAATAC 145
57 ..... 58
146 ATGAGTCTTTTTCACATCTTAATCGAATAATAGTAATTTGTAGAAATAT 195
|||||.....:.. |||||.....:..
58 IsLysGIuTyrlArgAsnIleIleAsnAsnAspSerLysAsnSerAsn 74
196 GGATCTTATTCGGGAGATGGGTGTAACTCTAGAAATTAACAAATAAATA 245
75 AsnSerPheCys.....AsnAsnGIuGI 82
246 AATAATAATAAATAATGAT.....AGAAATCACCAAGACAAATTTTATG 289
|||||.....:.. |||||.....:..
82 uAsnAsnLeuLysAspAspLeuLeuTyrlYsAsnSerArgLeuGIuAsn 99
290 AGGAGTATAAAAAAGAAACAAGTATTATATGCGAGTACTATAGTCTCA 339
|||||.....:.. |||||.....:..
99 LuAspPhe.....IleValThrGIyGIuValIlePhe 109
340 TGGAACTGTAAGGTGATAGAGCAAAACATATGATTGATCAAAACCAAT 389
|||||.....:.. |||||.....:..
110 AsnThrAlaMetValGIyTyrlProGIuAlaLeuThrAspProSerTyrlPh 126
390 GGTGTCAATTTTATATATTCATTTGCTCGCATTAATATGTTATATGATG 439
|||||.....:.. |||||.....:..
126 eGIyGIuIleLeuValLeuThrPheProSerIleGIyAsn...TyrgIyA 142
440 TATCTAGC.....CCATTAAATGGAAGCAAAAGATTCTCA 474
|||||.....:.. |||||.....:..
142 IeGIuLysValLysHisAspGIuThrPheGIuLeuValGIuAsnPhGIu 158
475 TTAGAAAA.....CACGGCTTA..... 492
159 SerAsnLysIleGIuValGIuGIuLeuValIleCysGIuTyrlSerLysGI 175
|||||.....:.. |||||.....:..
493 .....GAATATGAACCTATGATGATGATCTTAATGAATTAAGACGTA 535
|||||.....:.. |||||.....:..
175 uSerTyrlHisTyrlAsnSerTyrl...IleThrLeuSerGIuTyrlPheLys 191
536 TCACAAAAAGTA.....CGTCCAGATGTAAT 561
|||||.....:.. |||||.....:..
191 IeTyrlLysIleProCysIleGIyGIuIleAspThrArgAlaLeuThrLys 207
562 ATTGTTTATTCCTAGGTGAGAAACCTATATGATAGATAGAAAGA 611
|||||.....:.. |||||.....:..
208 LeuLeuArgGIuLysGIySerMetLeuGIyLysIleValIleTyrlLysAs 224

```





seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:us-08-405-496A-10  
2062 easnashnlys 2065

seq\_documentation\_block:  
; Sequence 10, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-405-496A-10

alignment\_scores:  
Quality: 114.50 Length: 654  
Ratio: 0.400 Gaps: 40  
Percent Similarity: 43.731 Percent Identity: 20.183

alignment\_block:  
US-09-579-383-2 x US-08-405-496A-10 ..

Align seg 1/1 to: US-08-405-496A-10 from: 1 to: 2366

33 AGATTCACATCTT.....GTATTCGCAATTCACGAA 64  
||||:|||||||  
1554 SerValHisIleuaspGluSerGlyValalaGluIleuLysPheMetAs 1570  
65 CCTTGAAGGAAAAATATATATATATATATTCATTGGGATATACGGGAA 114  
|||||||:||||| :|||||||  
|||||||:||||| :|||||||

1570 n.ArgIysGlyAsnThrAsnThrSerAspSerLeuMetSerPheLeu 1586  
115 AATATAAAATATAAATCATCAACGAAATACATGATGCTTTTCACATCT 164  
::: |||  
1587 SerMetAsn.....IleLysSerIlePheValAsnPh 1597  
165 TAAATGCAATTAATAGTAAATTTGTAGAAATATGATCTTATTCGGAGAT 214  
: :|||:||||| :|||:|  
1597 eleuGlnSerAsnIleLysPheIleuAspIleAsnPhelIleIleSerG 1614  
215 GGTGTAACCTGTAGATTAACAAAAATATAAATATAAATATAATAT 264  
|| :||| || :||| :|||:|  
1614 LysThrSer...IleGlyGlnPheGluPheIleCysAspIleuAsnAsp 1629  
265 AGAAATACCAAGACAAATTTTAGAGGATATAAAAAAGAAACAAG 314  
::: |||  
1630 AsnIleGlnPro..... 1633  
315 TATTAATAGCAGATACATATGTTTCATGGAACAGTCAGTGATAGACGA 364  
||||: |||:|  
1634 .....TyrPheIleLysPheAsnThr..... 1640  
365 AACATATGATTTGATTCACAAACCAATGCTGCATTTTATATATTCATTT 414  
: |||:|  
1641 .....LeuGluThrAsn.....TyrThrLeuTyrValGly... 1650  
415 GCTCGCATTAATATGTTA.....ATGATGTA..... 441  
||| |||:|  
1651 AsnArgGlnAsnMetIleValGluProAsnTyrAspLeuAspSerG 1667  
442 .....TCTAGACATTTAATGGAAGACAAAGATTCCTATTAGAA 481  
||| :||| :|||:|  
1667 YaspIleSerSerThrValIleAsnPheserGlnLysTyrLeu..... 1681  
482 AACACGGCTTAGAA.....TAT 498  
1682 ..TyrIleLysAspSerCysValAsnLysValIleSerProAsnIle 1697  
499 GAACCTATGATGATGATGCTTAATGAATATGACATACAGAAAGTACG 548  
||| :||| :|||:|  
1698 TyrThrAspGluIleAsnIleThrProValTyrGluThrAsnAsnThrT 1714  
549 TCCAGATGTAATTAATCTTTTATCTTATCTTATGAGTACATATGATAG 598  
|||:|  
1714 rProGluValIleValLeu..... 1720  
599 AATATGAAAAAGAAATTCATTTATGTCG...GATTAATATTCAGCTTGT 645  
1721 .....AspAlaAsnTyrIleAsnGluLysIleAsnValAsnIle 1733  
646 AATGATTTTGTATTTAGATGCTGATGATATTCAGTGGACCAATACGGA 695  
|||:|  
1734 AsnAspLeuSerIleArgTyrVal.....TyrSerAsnAspGlyAs 1747  
696 GTTTACAACTTA.....AATGAATTAATTTTCAATTAATTT 733  
: :||| :||| :|||:|  
1747 nasPheIleLeuMetSerThrSerGluGluAsnLysValSerGlnVal 1764  
734 ATATTAATTAATTAATTAATCTTTTANA...AAACATTCGCGAAGAAAG 780  
|||:|  
1764 ySileArgPheValAsnValPheLysAspLysThrLeuAlaAsnLysLeu 1780  
781 TTAATTTCAATTTCTGTTCAATCAAAATCTCATTAATCA.....TCGCT 824  
: |||:|  
1781 SerPheAsnPheserAspLysGlnAspValProValSerGluIleIleLe 1797  
825 TTCAGGAGTTCATCTTTCTGTAAGATGA..... 855  
1797 userPheThrProSerTyrTyrGluAspGlyLeuIleGlyTyrAspLeuG 1814  
856 .....GAATTCATATTAACACTAATTTTGTCTGAAACAATATGAAACA 900  
||| |||:|  
1814 LysLeuValSerLeuTyrAsnGluLysPheTyr..... 1824

```

seq_name: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-08-915-136-10
seq_documentation_block:
  Sequence 10, Application US/08915136
  Patent No. 6290960
  GENERAL INFORMATION:
    APPLICANT: KINK, JOHN A.
    APPLICANT: THALLEY, BRUCE S.
    APPLICANT: PADHAYE, NISHA V.
    APPLICANT: FIRCA, JOSEPH R.
    APPLICANT: STAFORD, DOUGLAS C.
    TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
    TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
    NUMBER OF SEQUENCES: 32
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: MELEN & CARROLL, LLP
      STREET: 220 MONTGOMERY STREET, SUITE 2200
      CITY: SAN FRANCISCO
      STATE: CALIFORNIA
      COUNTRY: UNITED STATES OF AMERICA
      ZIP: 94104
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentln Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/915,136
      FILING DATE:
        CLASSIFICATION:
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 08/480,604
        FILING DATE:
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/405,496
        FILING DATE: 16-MAR-1995
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/329,154
        FILING DATE: 25-OCT-1994
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/161,907
        FILING DATE: 02-DEC-1993
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/985,321
        FILING DATE: 04-DEC-1992
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/429,791
        FILING DATE: 31-OCT-1989
      ATTORNEY/AGENT INFORMATION:
        NAME: INGOLIA, DIANE E.
        REGISTRATION NUMBER: 40,027
        REFERENCE/DOCKET NUMBER: OPHD-01763
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: (415) 705-8410
          TELEFAX: (415) 397-8338
        INFORMATION FOR SEQ ID NO. 10:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 2366 amino acids
            TYPE: amino acid
            TOPOLOGY: linear
            MOLECULE TYPE: protein
            US-08-915-136-10

```

Quality: 114.50 Length: 654  
 Ratio: 0.400 Gaps: 40  
 Percent Similarity: 43.731 Percent Identity: 20.183

alignment block:  
 US-09-579-383-2 x US-08-915-136-10 ..

Align seg 1/1 to: US-08-915-136-10 from: 1 to: 2366

```

33 AGATCCATCTT.....GATTCGCAATTCACAA 64
   |||:|||||
1554 SerValHisLeuAspGluSerGlyValAlaGluIleLeuLysPheMetAs 1570
   - - - - -
65 CCTGGAAGAAAAATATATATAATTAATTCAGTAATACGGAA 114
   |||:|||||
1570 n.ArgLysGlyAsnThrAsnThrSerAspSerLeuMetSerPheLeuGlu 1586
   - - - - -
115 AATAAATATAACTCATCAACGCAATCATGATCTTTTCACACT 164
   |||:|||||
1587 SerMetAsn.....IleLysSerIlePheValAsnPh 1597
   - - - - -
165 TAAATCGAATAATAGTATTTGTAGAATATGATCTTATTCGGAGATG 214
   |||:|||||
1597 eleuGlnSerAsnIleLysPheIleLeuAspAlaAsnPhelIleSerG 1614
   - - - - -
215 GGTGTACTCTAGAAATTACAAAAATATAATAATTAATAAATGAT 264
   |||:|||||
1614 LyrThrThrSer...IleGlyGlnPheGluPheIleCysAspGluAsnAsp 1629
   - - - - -
265 AGAAATACCAACAGCAAAATTTTAGAGAGATATAAAAAAGAAACAAG 314
   |||:|||||
1630 AsnIleGlnPro..... 1633
   - - - - -
315 TATTATAGCAGGATATGATGTTGATGAGACAGTCACAGTGATAGACAA 364
   |||:|||||
1634 .....TyrPheIleLysPheAsnThr..... 1640
   - - - - -
365 AACATATGATTTGATCAACCCCAATGGTGCAATTTTATATATTCATT 414
   |||:|||||
1641 .....LeuGluThrAsn.....TyrThrLeuTyrValGly... 1650
   - - - - -
415 GCTGCATTAATATGTTA.....TATGATGTA..... 441
   |||:|||||
1651 AsnArgGlnAsnMetIleValGluProAsnTyrIleAsnAspAspSerG 1667
   - - - - -
442 .....TCTAGACCAATTAATGAGAACCAAGATTCCTATTAGAA 481
   |||:|||||
1667 yAspIleSerSerThrValIleAsnPhSerGlnLysTyrLeu..... 1681
   - - - - -
482 AACACGGCTTAGAA.....TAT 498
   |||:|||||
1682 ..TyrGlyIleAspSerCysValAsnLysValAlaIleSerProAsnIle 1697
   - - - - -
499 GAACCAATGATGATGCTTAATGAATAGAGTATCAGAAAGATGACG 548
   |||:|||||
1698 TyrThrAspGluIleAsnIleThrProValTyrGluThrAsnAsnThrTy 1714
   - - - - -
549 TCCAGATGTAATATCTTTATCTTACCTAGGTGAGAAACCTATATGATAG 598
   |||:|||||
1714 rProGluValIleValLeu..... 1720
   - - - - -
599 ATATTACAAAAACAATATGATTATGTG...GATAAAATATTGCAACTGTT 645
   |||:|||||
1721 .....AspAlaAsnTyrIleAsnGluLysIleAsnValAsnIle 1733
   - - - - -
646 AATGATTTTGATTTAGATGCTGTAGATATGACGGGAAACCACTGGAA 695
   |||:|||||
1734 AsnAspLeuSerIleArgTyrVal.....TyrSerAsnAspGlyAs 1747
   - - - - -
696 GTTTTCAACTTA.....AATGAAATTAATTTTCAATATATT 733
   |||:|||||
1747 nsppheIleLeuMetSerThrSerGluGluAsnLysValSerGlnVal 1764

```

```

734 ATATTAAATTAATTAACCTGTGTAGA...AAACATTTCCGAGAAAG 780
   |||:|||||
1764 ySleIleArgPheValAsnValPheLysAspLysThrLeuAlaAsnLysLeu 1780
   - - - - -
781 TTAATTTCAATTTCTGTTCAATCAATGCTGCATTATCA.....TGCCT 824
   |||:|||||
1781 SerPheAsnPheserAspLysGlnAspValProValSerGluIleIle 1797
   - - - - -
825 TTCAGAGAGTTCGATCTTCTGTAAAGATGAA..... 855
   |||:|||||
1797 uSerPheThrProSerTyrTyrGluAspGlyLeuIleGlyTyrAspLeuG 1814
   - - - - -
856 ....GAATCTCCATATACACTAAATTTTGTCTGAACAAATAGAAACA 900
   |||:|||||
1814 lYleuValSerLeuTyrAsnGluLysPheTyr..... 1824
   - - - - -
901 AATAAGAAATTAACATAGGCGACGACGATGTATACAGCAGCACTTTAT 950
   |||:|||||
1825 .....IleAsnAsnPhelGlyMetValSerGlyLeuIleTyrIl 1838
   - - - - -
951 TAAATTTTAAATACACCAAGAGAAAAATAGATCTGTATTATTTCAAA 1000
   |||:|||||
1838 eAsn.....AspSerLeuTyr..... 1843
   - - - - -
1001 CATACAAATTAGAACTACAAATCCAGATATATAGTATGATGACTTA 1050
   - - - - -
1843 ..... 1843
   - - - - -
1051 TCCATTTATATTTGCTTAAATATATACATCAATCATATAGGTTT 1100
   |||:|||||
1844 .....TyrPheLysProProValAsn...AsnLeuIleThrGlyPh 1856
   - - - - -
1101 T...TCATTAGAACATACACAGGCTGATTTAGTCCGAAATTAAGAAAT 1147
   |||:|||||
1856 eValThrValGlyAspAspLysTyrTyrPheAsnProIleAsnGlyLys 1873
   - - - - -
1148 TATTAGAAATTTGCTAGGAAAAACAATFACATGATAAAATCAATAATAT 1197
   |||:|||||
1873 laaIleSerIleGlyGluThrIleIleAspLysAsnTyrTyrPheAsn 1889
   - - - - -
1198 AGGCACAGATGATAGGATATGTCATTATTATGAAACAACTATACC 1247
   |||:|||||
1890 GlnSer.....GlyVal.....LeuG 1895
   - - - - -
1248 AACTGATCATTCGATGATAGATATTTTCTACAAATATTTGGAACATT 1297
   |||:|||||
1895 nThrGlyValPheSerThrGlu.....AspGlyPheLysTyrP 1908
   - - - - -
1298 TAAATCTGGAAGTACAAACTCCAAAGACCTTACTACTGAAACCT 1347
   |||:|||||
1908 heaIaPro..... 1910
   - - - - -
1348 GAACACGTGTAGCACATAGATATATGTTCCAGAGCTGTATTCACAC 1397
   |||:|||||
1911 .....AlaAsnThrLeuAspGluAsnLeuGlnGlyAlaIleAspPh 1925
   - - - - -
1398 CATAGG.....ATAATATAC.....AAAC 1417
   |||:|||||
1925 eThrGlyLysLeuIleIleAspGluAsnIleTyrTyrPheAspAspAsn 1942
   - - - - -
1418 ACAATGATGCTATA...TGAAA..... 1437
   |||:|||||
1942 yTArgGlyAlaValGluTyrPlyGluLeuAspGlyGluMetHisTyrPhe 1958
   - - - - -
1438 .....ACTAG 1442
   - - - - -
1959 SerProGluThrGlyLysAlaPheLysGlyLeuAsnGlnIleGlyAspTy 1975
   - - - - -
1443 ATCTTATTCATTCATGACCTGTGTGA..... 1470
   |||:|||||
1975 rLysTyrTyrPheAsnSerAspGlyValMetGlnLysGlyPheValSerI 1992
   - - - - -
1471 .....GACAGATATGAATGGGAC.....TTGGTCAAAATATGTC 1503

```

1008 yaspGluLysAlaPheAspAsnAlaLysLysAspLeuGlnSerLeuLeuG 1025  
191 AATATGATCTTATTGCGGA.....GATGGGTGAACCTCAGAAATT 231  
1025 LuThrTyrSerLysSerAspSerSerIleAspValPheLysMetSerPhe 1041  
232 ACAAAAATAATATAAATAATATAATATAATATGATAAATACCAACAGCA 281  
1042 AspLysAlaGlnLysAsnIleLysAspPolysp..LysSerLeuSerSe 1057  
282 AATTTTAGAGAGAT..... 297  
1057 rValLysSerGluValGlyAspLeuGlyLuThrLeuAlaGluAlaGlyA 1074  
298 .....AAAAAGCAAAACGATATTATAGCAAGA 327  
1074 sngLuAlaGluAspPheGlyLysLysLeuLysGluValAlaLeuAspAla... 1089  
328 TACTATGGTTCATGAGACAGTCACAGTGAT.....AGACCAACA 368  
1090 .....AsnSerValAspSerPheLysAlaIleLysGln 1101  
369 TATGATTTGATCAAAACCAACGGTGCTCAATTTATATATTCATTGCTC 418  
1101 MetSerAspAlaMetGlnPheAspSerValGlnAspValLeuAsnGlyA 1118  
419 GCATT...AATATGTTATATGATGATCATCAACCATTTAATGCAAGCAAA 465  
1118 sPLePheAsnAsnThrLysAspGlnValAlaProLeuAsnAsp..... 1132  
466 ACATTCTTATTAGAAAACACGGCTTAGAATATGAAACCTATGGTATGAT 515  
1133 .....LeuLeuGluLysMetAla.....GluGlySerIleSe 1144  
516 GCTTATGAAATATGACCTATACAGAAAGATACGTCACAGATGATTATTC 565  
1144 rAlaAsnGluAlaAsnThrLeu.....IleGlnLysAspLysGluLeuA 1159  
566 TTTTATCTTAGGTGGAGAACCTATATGATAGATAGAAAAAGAA... 612  
1159 IaGlnAlaIleSerIleGluAsnGlyValValLysIleAsnArgAspGlu 1175  
613 .....ATTGA 617  
1176 ValIleLysGlnAlaGlyLysValLysLeuAspAlaTyrAsnAspMetValTh 1192  
618 TTATGTGATATAAATATTGAAG.....CTTGTATATATT 652  
1192 rTyrSerAsnLysLeuMetLysThrGluValAlaAsnAlaIleLysThrL 1209  
653 TTGATTATGATGGTGCTAGATATTGAC.....TGGGAACCAT 690  
1209 euAsnAlaAspThrLeuArgIleAspSerLeuLysLysLeuArgLysGlu 1225  
691 GGGAGCTTTTCAACTTAAATGAATATTAATTTTCAATATATTATTATAA 740  
1226 ArgLysLeu...AspMetSerGluAlaGluLeuSerAspLeuGluValLy 1241  
741 ATTAAATCACTTGTA.....AGAAAACTATTCCG.....GAAG 775  
1241 sSerIleAsnAsnValAlaAspAlaLysLysGluLeuLysLysLeuGln 1258  
776 AAAACTTAAATTCATTTCGGTTCATCAACAATGGTCATTATCATGGCTT 825  
1258 LuLysMetLeuGlnProGlyLysTyrSerAsnSerGlnIleGluAlaMet 1274  
826 TCAGAGATTCGATCTTCTGTAAGATGAGAGAATCTGCATATAACACTAA 875  
1275 GlnSerValLysSerAlaLeuGlnUserTyrIleSer..... 1286  
876 ATTTTGTCTGACACAAATAGAAACAAATAAAGATTACATAGGCGACAG 925  
1287 ...AlaSerGlnGluAlaThrSerPheGlnGluMetAsnLys...GlnA 1301

```

926 CGATGTTATTCAGCAGAACTTTTATTAATATTTTATATACGACGAAGAG 975
1301 IAEUVALGUAAGLYTHSerLeuGluGlnSerPheASPGLnGlnIleu 1317
976 AAAATGATCTGTATTATTTCAACAATACATTTAGAAACTACAAAATCC 10259
1318 LysAla.....AsnLugluThrLysThS 1320
1026 AGATATAATGCTGATATGTACTTATCCCATTTATATTTGGTTAAAT 10757
1326 rMetYrValValAspLysTyr..... 13333
1076 ATAAATCACAATATCATTTAGTTTTCATTAGAACATACAGAGGTGGA 11255
1333 ..... 13333
1126 TTTATGCCGAAATAAAGATTTATTAAGATTGCTAGAAAAACAATACA 11757
1334 .....LysGluAlaLeuGluLysValAsnAlaGluIleAs 1345
1176 TGATAAAATCAAAATATATAT 1197
1345 pLysTyrAsnLysGluValAsn 1352
seq_name: /cgn2_6/ptodata/2/laa/6A_COMB.pep:US-08-949-588-2
seq_documentation_block:
; Sequence 2, Application US/08949588
; Patent No. 6025156
; GENERAL INFORMATION:
; APPLICANT: Gwynn, Michael
; APPLICANT: Kallendar, Howard
; APPLICANT: Palmer, Leslie
; TITLE OF INVENTION: No. 6025156e1 Topoisomerase III
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/949,588
; FILING DATE: 14-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/028,417
; FILING DATE: 15-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50567
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-949-588-2

```

```
alignment_scores:  
    Quality: 112.50  
    Ratio: 0.434  
Percent Similarity: 44.965  
  
Length: 576  
Gaps: 31  
Percent Identity: 20.833
```

```
alignment_block:
US-09-579-383-2 x US-08-949-588-2 . .
```

Align seg 1/1 to: US-08-949-588-2 from: 1 to: 711

[illegible]



```

949 .....ATTAAATATTTTAAATACAGCAAG..... 972
347 rlymetserillepheasnansglnlysvalseasphishtalalei 364
973 .....GAGAAATAGACTCTTATTTATTCACAAATACATATTGAAACT 1017
364 leprothrgluvalargprovalmetseraspleuserasnarglndeu 380
1018 ACAATCCAGATATATAGTATGATCTACTATCC..... 1053
381 lysleutryraspmetillevalgluarphleuclualaleuemetprop 397
1054 .CATTATAT.....T 1063
397 ohisglutryaspalallerhrvalthrleugluvalalaglysthrp 414
1064 TTGGTTTAAATATACATCACATCATATAGTTTTTCATTGAAACAT 1113
414 hevalleulysgluasvalthrvalleuglypne..... 426
1114 AACAGAGTGATTTAGTCCCGAAATTAAGAAATTAATGAAATGTTAG 1163
427 .....lyserillearggin.....GI 432
1164 AAAACAATCATGATATAAAATCAAAATTAATTAAGGCGAGATGTTAG 1213
432 ygluserlthrglumetgnglnpropheserlglglyasp..... 446
1214 GGATATGSCATTTTATTAAGAAACAATACCACTGGATCATTCGAT 1263
447 .....glu 447
1264 GTAGATATTTTCTTCAAAATATTGGAACATTAAATCTGAAGTACA 1313
448 valylsileserysthrasnilleargluhts.....gluth 460
1314 AACGCCAAAAGAC.....CTTACTATAACTGAAA 1342
460 rthpproglnutryrphasnngllyserleulesslameglua 477
1343 ACCCTGAAGACTGAGCACAATA...GATGAATATGTTCCAGACGCTGT 1389
477 snproglnasnphelleglndeuylsaspys..lystfyracglnthrle 493
1390 ATTCCAACCATAGGATATATTCAAAACAAATGATGCTATATGAAAAC 1439
493 ulysglntthrlyllylellythr.....Valalathrarfga 506
1440 TAGATCTTATTCATTCATGCACCTGGTGTAGACAGATATGATGGACT 1489
506 laaspillelleasplysleuphe.....Asnmetasnalaile 518
1490 TGGCAAGATAGTCT...ATGAAAAAATATGCGATGGGAAGCGACCCAT 1536
519 glusertraspglyllysllysvallthrseryscllysglnllele 535
1537 T.....ATTATAACTGACTATAA 1556
535 ugluleuAlaproglucluleuthrserproleucluleuthralaglntrpg 552
1557 AGAAAGCTCTATTTATATGGAAGGGAACATATTATTAATAAGT 1606
552 lugalylsleucluleu..ilegluargly..lystfyrclnalalysthr 568
1607 GGCAACAAGAGACCTCCGGAAGGTGAGSCACATACATACACAACACTA 1656
568 helle.....Asnclumetlysasphethlysh... 577
1657 GATCATCAATATGTCCAGGATAGAAATGAAATATAATATATCCACA 1706
578 .....Aspvalvalasnnglyllylelysnasnsersparglystfyrlyshi 592

```

```

1707 TAAACACTAGAGTAGAGGAA 1728
592 saspsanleuthrthrthrthr 599
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-912-4
seq_documentation_block:
; Sequence 4, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: leary, kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1861 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-4

alignment_scores:
Quality: 109.00 Length: 537
Ratio: 0.441 Gaps: 27
Percent Similarity: 45.996 Percent Identity: 20.112

alignment_block:
US-09-579-383-2 x US-08-790-912-4
Align seg 1/1 to: US-08-790-912-4 from: 1 to: 1861
76 AAAAATATAATAATAAT.....TCATTGGGAATAATA..... 108
874 Glucllysnlleasnansvallysarphneclyservalalaglyvalal 890
109 .....CGGAAATTAATAATTAACATCAACAAAGGAAA 142
890 aglytyrleutrpaspargspserserglugluarghislalaglyr 907
143 TACATGAGTCTTTTCAACATCTTAATCGAATATAGTAATTTGTGAA 192
907 euhisasnvalleuserasplleasnvalmetasnnglyasnalaileser 923

```



```

? FILING DATE: 08-DEC-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Stone, Paul A.
? REGISTRATION NUMBER: 38,628
? REFERENCE/DOCKET NUMBER: SLU 4471
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 314-231-5400
? TELEFAX: 314-231-4342
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 223 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM: Human parainfluenza virus 3
? STRAIN: JS
? US-08-569-853-1

```

```

alignment_scores:
    Quality: 109.00      Length: 673
    Ratio: 0.355        Gaps: 32
    Percent Similarity: 45.617    Percent Identity: 18.425

```

```

alignment_block:
US-09-579-383-2 x US-08-569-853-1 ..

```

```

Align seg 1/1 to: US-08-569-853-1 from: 1 to: 2233

```

```

4 AATTTTAAATATCAATATTTTATTAATAGTACCATCTGTATTCGC 53
||||| : : : : : ||| : : : : : ||| :
1270 AsnPhetHrLeuSerLeuLysIleLeuThrProValAlaThrSerTh 1286
54 AAT...TCCAGAACCTGGAAGA..... 75
||| ||| : : : : : |||
1286 rAsnLeuSerHsArgLeuLysAspThrAlaThrGlnMetLysPheSer 1303
76 .....AAAAATATTAAT 90
1303 erThSerLeuIleArgValSerArgPheIleThrMetSerHsAspAsn 1319
91 AATTCATTGGGAATATACGGGAAATAA.....AATAAACTCATCA 134
||| : : : : : ||| : : : : : |||
1320 MetSerIleLysGluAlaAsnGluThrLysAspThrAsnLeuIleTyrG 1336
135 A.....ACGGAATACATGAGCTTTTCACATCTTA 166
| : : : : : ||| : : : : : |||
1336 nGlnIleMetLeuThrGlyLeuSerValPheGluThrLeuPheArgLeu 1353
167 AA...TCGAATATATAGTATTTTGTAGAATATGATCTTATTCGGGAT 213
|| : : : : : ||| : : : : : |||
1353 YsgIuThrThGlyHsAsnProIleValMetHsLsLeuHsIleGluasp 1369
214 GGGGTAACTAGTAATTAACAAAAATATAAATATAAATAAATAAAGA 263
|| : : : : : ||| : : : : : |||
1370 GlucysysIleLysGluSerPheAsnspGluHsIleAsn..... 1383
264 TAGAAATACACCAAGACAAATTTTAGAG.....GAGTATAAAAAAGA 307
||| : : : : : ||| : : : : : |||
1384 .....ProGluSerThrLeuGluLeuIleArgTyrProGluSerGAT 1397
308 AACAAAGTATATAGACAGATAGTATGTCATGGAACAGTCAAGGTGAT 357
: : : : : ||| : : : : : |||
1397 snGluPheIle.....TyrAspLysAspProLeuLysAspValAsp 1410
358 AGAGCAAAACATATG.....ATTGATTCAAA 383
: : : : : ||| : : : : : |||
1411 LeuSerLysLeuMetValIleLysAspHisSerLysThrIleAspMetAs 1427
384 CCCAATGGTGTCAATTTATATATGCAATTGTCGCCATTAATATGTAT 433

```

```

1427 n.TyrTrpAspSphrAspIleIleHsAlaIleSerIleCysThrAla 1443
||| : : : : : ||| : : : : : |||
424 ATGATGATCTAGACCATTAATGGA...GACAAAGATTCCTATTAGA 480
||| : : : : : ||| : : : : : |||
1444 IleThrIleAlaAspThrMetSerGlnLeuAspArgAsp..... 1456
481 AAACAGCGCTTAGAATATGAACCTATGATGATGATGATGATGATGAT 530
||| : : : : : ||| : : : : : |||
1457 .....AsnLeuLysGluIleIleValIleAlaAsnAspAspA 1469
531 ACGTATCAGAAAAGTACGCCAGAT.....GTAATTAT 564
||| : : : : : ||| : : : : : |||
1469 spLsAsnSerLeuIleThrGluPheLeuThrLeuAspIleLeuValPhe 1485
565 CTTTATACCTTAGTGTGAGAAAACCTATATGATATAGTAAAGAAAT 614
||| : : : : : ||| : : : : : |||
1486 LeuYsthrPheGly..... 1491
615 TGATTTATGTGATAAATATTTGAAGCTTGTATATGATTTT..... 654
||| : : : : : ||| : : : : : |||
1492 .....LeuLeuValAsnGlnPheAlaTyrThrL 1501
655 .....GATTAGATGCTGATAGATATGACTGGGAACCATGGG 693
||| : : : : : ||| : : : : : |||
1501 euTyrSerLeuLysIleGluGlyArgAspLeuIleTrpAsp..... 1514
694 AAGTTTACAACTTAATGAATTAATTTTCAATTTATATATTAATTT 743
||| : : : : : ||| : : : : : |||
1515 .....TyrIleMetArgThrLeuArgAspThrSerHis.....SerI 1527
744 AATTAACCTGTTAAGAAAACCTATTCGGAAGAAAGTAAATTTCA.... 789
||| : : : : : ||| : : : : : |||
1527 eleuLysValLeuSerAsnAlaLeuSerHisProLysValPheLysArg 1544
790 .....ATTTCGTGTCATCAATGCTGCA 813
||| : : : : : ||| : : : : : |||
1544 heTrpAspCysGlyValLeuAsnProIleTyrGlyProAsnThrAlaSer 1560
814 TTATCATCGTTTCAGAGTCATCTTCTGTAAGATGAAGATCTCC 863
||| : : : : : ||| : : : : : |||
1561 GlnAspGlnIleLysLeuAlaLeuSerIleCysGluTyrSerLeuAspLe 1577
864 ATATTAACACTAAATTTTGTCT..... 885
||| : : : : : ||| : : : : : |||
1577 upHemetaArgLutPrLeuAsnGlyValSerLeuGluIleTyrIleCysA 1594
886 .....GAACAATAGAAAACAATATAA 906
||| : : : : : ||| : : : : : |||
1594 spSerAspMetGluValAlaAsnAspArgLysGlnAlaPheIleSerArg 1610
907 GAATTAACATAGGCGAGCGAGTGTATCAGCAGGACCTTT...ATTAA 953
||| : : : : : ||| : : : : : |||
1611 HisLeuSerPheValCysCysLeuAlaGluIleAlaSerPheGlyProAs 1627
954 TATTTTAAAT...ACAGCAAAAGAGAAATAAGATCTGTATTTATTCAAA 1000
||| : : : : : ||| : : : : : |||
1627 nLeuLeuAsnLeuThrTyrLeuGluArgLeuAspLeu.....LeuLysG 1642
1001 CATAC.....AATTTAGAACTACAAATCCGATATATATGTTAGAT 1041
||| : : : : : ||| : : : : : |||
1642 InTyrLeuGluLeuAsnIleLys...GluAspProThrLeuLysTyrVal 1657
1042 ATGTACTATATCCATTTATTTGTTTAAATATTAATATACATACATCAT 1091
||| : : : : : ||| : : : : : |||
1658 GlnIleSerGlyLeuLeuIleLysSerPheProSerThrValThrTyrVa 1674
1092 ATTAGGTTTTCATTAGAACATACAGA...GGTGGATTTAGTCC... 1134
||| : : : : : ||| : : : : : |||
1674 ArgLysThrAlaIleLysTyrLeuArgIleArgGlyLsSerProPog 1691
1135 .....GAAATTAAGAATTTTGAAGATTG 1158
||| : : : : : ||| : : : : : |||

```

1691 luvalillleaspsaprrlpasprovalgluaspolusnmetleuaspsn 1707  
1159 GTAGCAAAAACAATACATGATATAAAATCAAAATATAATAGGGCAGATGG 1208  
::: |||||||:::||||: :::::||||: :::::  
1708 llevallystrhilleasnpspsncysasnlysaspsnlysglsyanly 1724  
1209 TATAGGGATATGGCAATTATATTATGAAACAATACCACATCGATCAT 1258  
|||  
1724 site..... 1725  
1259 TCGATGTGATATATTTCTTACAAATATTGGAAACATTTAAATCCGAA 1308  
:::|||||:::||||: |||  
1726 .....AsnAsnphetiRgyleualaleuLysAsn 1735  
1309 GTACAAACTCCAAAAGACCTTACTATACGTGAAACCCTGAAGACGTAG 1358  
||||: ||| :::|||||::: ::::: |||  
1736 TyrglnvalleuLyslleargserlIethrSerAspspsnAs 1752  
1359 CACAAATAGATAAATATGTTCCAGGACTGTTATCCAACCATRAGCATAT 1408  
::: ||| :::: |||||||::: ||| |||  
1752 pargleuAspsllaSnthrSerGlyleuThrleuprogInglylAsnT 1769  
1409 ATTCAACAACCAATATGCTATATGCAAAATAGATCTATTACA..... 1452  
|| :::|||||::: ::::: |||  
1769 yrleuSerHnSgInleuargleupheglylleasnerhrSerCysleu 1785  
1453 .....ATTCAATGCACCTGTGTAGACAGATATGA 1481  
1786 lysalaleuInleuSerGlnleuemetLysglValasnLysAsply 1802  
1482 ATGGGACTTG.....GTCAAAAGTATGCTATG 1507  
::: ||| ::::: |||||||  
1802 saspargleuPheleuInglyglYalaglYalameLleuAlacysTryA 1819  
1508 AAAAATATGCGATGGAAAGCAGCCCATTTATATACACTACTATAAA 1557  
1819 spAlaThrleu...GlyProAlaValasnTryTyrsmSerGlyleuAsn 1834  
1558 GAAAGCTCTATTAT.....ATATGAAAGGGGAACC 1589  
:::|||||::: |||: ::|||  
1835 llehrspvalillecglylnargleuLysllePheProserGluva 1851  
1590 ATATTATATTAATATGGTGGCAACAAGACCTCCGGAAGTCAGCAGCATAG 1639  
|||::: |||  
1851 lserleuVal.....GlyLysleuG 1859  
1640 AGTCATACCAAAACTA 1656  
1859 lysanvalThrGlnle 1864

OM of: US-09-579-383-2 to: PIR-68.\* out\_format : pfs

Date: Mar 21, 2002 4:42 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=frame+np2.model -DEV=xlp
-O=cpq2.1/USPTO.spool/US09579383/runtat_20032002.151239_29320/app-query.fasta_1.1849
-DB=PIR-68 -PEMT=fastan -SUFFIX=PIR -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blonsun62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pcit -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09579383_@CGN1_1_89 -NCPU=6 -ICPU=3
-LOGLOG -NO_XLPHY -WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-579-383-2
Query length: 1764
Database: PIR-68.*
Database sequences: 219241
Database length: 76174352
Search time (sec): 99.360000
```

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
PIR1:S34960	- 158.50	184.71	0.0045	590	NADH dehydrogenase (ubiquinone)
PIR2:T18482	- 150.50	163.60	0.0155	2437	hypothetical protein C0705C -
PIR2:S61166	- 149.00	174.61	0.0191	511	probable membrane protein YDR37
PIR2:S72284	- 148.50	168.86	0.0213	960	DNA-directed RNA polymerase (EC
PIR2:T43916	- 148.00	166.33	0.0233	1215	chitinase A (imported) - Pyroc
PIR2:S60651	- 145.50	169.67	0.0328	563	chitinase precursor - Serratia
PIR2:A33221	- 145.00	169.98	0.0352	504	chitinase (EC 3.2.1.14) MFI - n
PIR2:H71611	- 144.00	168.29	0.0437	1817	probable secreted protein PFB0
PIR2:T24076	- 143.00	168.64	0.0474	444	hypothetical protein R09D1.5 -
PIR1:I130010	- 142.50	167.04	0.0514	502	NADH dehydrogenase (ubiquinone)
PIR2:A25090	- 142.50	166.13	0.0517	561	chitinase (EC 3.2.1.14) - Serrat
PIR2:B71621	- 139.50	161.04	0.0876	2295	probable membrane associated p
PIR2:T30933	- 138.50	156.23	0.0981	1054	chitinase (EC 3.2.1.14) A - p
PIR2:B71612	- 138.00	144.77	0.1132	3973	hypothetical protein PFB0555c
PIR1:O007C5	- 136.50	158.60	0.1292	590	NADH dehydrogenase (ubiquinone)
PIR2:T03884	- 136.50	148.51	0.1375	2025	hypothetical protein P07G11.9
PIR2:S07915	- 136.00	152.57	0.1442	1146	RF2 protein - yeast (Kluverom
PIR2:T41863	- 135.50	157.96	0.1500	552	chitinase chi-A orf126 - Bombyx
PIR2:G71609	- 135.50	145.59	0.1618	2500	hypothetical protein PFB0650w
PIR2:B26696	- 135.00	159.17	0.1601	443	hypothetical protein 1 (CYF-CO1
PIR2:T18460	- 133.50	152.19	0.2076	836	hypothetical protein C0510w - m
PIR2:T18497	- 133.50	142.72	0.2200	2657	hypothetical protein C0780w -
PIR2:T18501	- 133.50	140.72	0.2227	3394	hypothetical protein C0760c -
PIR2:T18414	- 133.00	140.82	0.2393	3119	hypothetical protein 9377 - malaria parasit
PIR2:S58360	- 132.50	149.81	0.2435	967	antibiotic Peps biosynthesis p
PIR2:A45624	- 131.50	152.97	0.2611	569	tropochoite cysteine proteinase
PIR2:A45597	- 131.00	140.80	0.3198	2339	DNA-directed RNA polymerase (B
PIR2:T29275	- 130.50	143.93	0.3373	1484	hypothetical protein T01C.1 -
PIR2:T28160	- 130.00	139.04	0.3737	2510	hypothetical protein - malaria
PIR2:T18440	- 130.00	134.17	0.3850	4550	hypothetical protein C0425w -
PIR1:A44267	- 129.50	151.32	0.3727	521	protein-tyrosine-phosphatase (B
PIR2:G72865	- 129.50	150.86	0.3737	551	chitinase - Autographa californi
PIR2:T24085	- 129.00	152.26	0.3984	432	hypothetical protein R09D1.11 -
PIR2:T18461	- 129.00	147.54	0.4100	768	hypothetical protein C0305c - m
PIR2:T18429	- 129.00	140.99	0.4268	1711	hypothetical protein C0345w -
PIR2:G70163	- 129.00	139.06	0.4319	2166	hypothetical protein B00512 -
PIR2:T00323	- 128.50	146.31	0.4442	831	chitinase (EC 3.2.1.14) B - Cld
PIR2:S41649	- 128.50	139.73	0.4624	1855	DNA polymerase - malaria paras
PIR2:T24074	- 128.00	151.01	0.4640	435	hypothetical protein R09D1.3 -
PIR2:T09079	- 127.50	135.45	0.5487	2708	probable chloroquine resistant
PIR2:T23643	- 127.00	149.04	0.5429	479	hypothetical protein M01B2.6 -

```
PIR2:E70118 + 126.50 145.76 0.5955 665 | flagellar hook-associated pr
PIR2:C71622 + 126.50 136.83 0.6290 1979 | hypothetical protein PFB014
PIR2:T18427 + 126.50 131.65 0.6492 3724 | hypothetical protein C0335c
PIR2:F82865 + 126.50 129.77 0.6567 4688 | hypothetical protein U0482
```

seq\_name: PIR1:S34960

#### seq\_documentation\_block:

```
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Crithidia oncopelti mitochondr
C:Species: mitochondrion Crithidia oncopelti
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C:Accession: S34960
R:Maslov, D.A.; Horvath, A.; Giang II, K.; Kolesnikov, A.A.
submitted to the EMBL Data Library, October 1990
A:Reference number: S34958
A:Accession: S34960
A:Molecule type: DNA
A:Residues: 1-590 <MAS>
A:Cross-references: EMBL:X56015; NID:g12879; PIDN:CAA39492.1; PID:g12882
C:Genetics:
A:Gene: ND5
A:Genome: mitochondrion
A:Genetic code: SGC6
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation
```

#### alignment\_scores:

```
Quality: 158.50 Length: 681
Ratio: 0.576 Gaps: 41
Percent Similarity: 40.382 Percent Identity: 21.145
```

#### alignment\_block:

US-09-579-383-2/rev x S34960 ..

Align seg 1/1 to: S34960 from: 1 to: 590

```
1743 TTTCTGTTTCATTTGCTCTCTACTCTAGTCGTTATGATTTT 1694
|||||:|||||:|||||: ||| :|||||: |||||
4 PhepPheMetLeuPhePheLeuPheGlyPhePheLeuGlyThPhePh 20
1693 T.....ATTTCATCTCTCATCC..... 1675
20 eleuGlyArgHisLeuSerPheThrPheSerIleLeuMetIValp 37
1674 .....TGACATTTGGATGCATCTTTTGTGTA 1645
37 heLeuValMetValThrMetPheSerPhePheCysIleSerValCysLeu 53
1644 TGACCTTAGTGC...CTGACCTTCGAGAGCTTGTCGCCACCATTTAA 1598
|||||
54 TyrGlyTyrCysTyrTyrAspPheCysLeuIleLeuMetLeuAspLeu 70
1597 TTAATTAATGATCCCTTCATATATATAGACCTTTTATAGTCA 1548
70 sPheIlePheMetSerPheTyrCysAsnGlyPheTyrIleuPheIleuP 87
1547 GTGTATATA...ATAATGGCTGCTTTCCATCGCATAT...TTTTCATA 1504
|||||:|||||: |||||: |||||
87 heLeuIleAspLeuValPheCysPheIleLeuPheTyrAlaPheTyrTyr 103
1503 GCATCTTTCACCAATGC.....CCATTCATATCTGTC.....TA 1469
|||||:|||||: |||||: |||||
104 MetTyrTyrAspLeuMetLeuAsnArgPhePheAsnIlePheTyrPPh 120
1468 CACCGAGGATCATTTGAATAGATCTAGTTTCCATATACCATCATTTG 1419
|||||:|||||: |||||: |||||
120 eValLeuCysMetAsnPheIleLeuSerTyrIspTyr..... 133
1418 TGTATTATATATATCCAT...GGTTGAAATACAGAGCTCGAACAATA 1372
|||||:|||||: |||||: |||||
134 .....LeuThrAlaTyrCysGlyTyr.....GluLeuLeuGlyLeu 145
```

```

1371 TTCAATCATGTGCTACAGTCTTCAGGGTTTCACTTATAGTAAGCTCTT 1322
1372 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1373 PheSerPhe.....PheLeuIleSerTyrPheTrpTyrArgph 158
1374 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1375 TTGAGCTTTGACTTCAGGATTTAATGTTTCCAAATATTTGTAGAGAAA 1272
1376 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1377 ePheAlaLeuLysPhe.GlyPheLysSerPhe.....PheIleSerLys 172
1378 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1379 ATATCATCATGGAATGATCCAGTGTGTAATGTTCTTCATTAATAAATG 1222
1380 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1381 IleGlyAspVal..LeuLeuLeuLeuSerPheValIleMetThrPheIleSer 189
1382 CC.....ATATCCCTATACCATCTGCGCCCTATATTATTATTGAT 1184
1383 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1384 hrgLysTyrGlyMetIleAsnPhetYrPheValAsnPhenLeuLysValAsp 205
1385 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1386 TTTTATCATGATGTTTCTTCTACCAATTCATTAATCTTATTATTTCG 1135
1387 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1388 PheTyrPheThrAlaPheMetLeuPheLeuIleMetCysAlaPheTh 222
1389 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1390 GGACTAAATCCACCTCTGTTATGTTTAAATGAAA 1100
1391 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1392 rlySerThrGlnPheGlyLeuHisIleTrpLeuProAspAlaMetGluG 239
1393 AACCT.....AATATGATGTTG... 1083
1394 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1395 LyrProIleProValSerAlaLeuIleHisAlaAlaThrLeuValValCys 255
1396 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1397 ATGTTT 1078
1398 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1399 GlyIleLeuLeuValSerPhePhePheTrpCysPheAspPheTrpLeuAl 272
1400 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1401 ATATTTTAAACCAAAATATAATAGGATAGACATATCATCATATATAT 1028
1402 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1403 atyPheTyrProLeuIleGlyTrpSerSerLeuIleLeuValMetMetS 289
1404 CTGATTTGCT.....AGTTTCT 1011
1405 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1406 er..LeuLysValPheTyrAsnPheAspAlaLysArgPheValAlaPheS 305
1407 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1408 AAATTTGATGTTTGAATATAATACAGATCATATTTCTCCTTCG... 967
1409 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1410 ertThrIleCysGlnIleSerPheSerMetPheCysCysLeuLysLeuAsp 321
1411 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1412 TGTATTAAAAATATTAATAAAGTCTCCTGATTAACA 929
1413 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1414 LeuTyrValIleGlyCysLeu..... 327
1415 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1416 TCGCTGCTGCCCTATGTAATCTTATTTGTTTATTTGTTTGTACAGACAA 879
1417 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1418 PhePheCysTyrHisMetPheTyrLysA 337
1419 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1420 TTTAGTGT.....ATATG... 862
1421 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1422 IatThrLeuPheIleValLeuGlyValItrPleHisLeuPhePheGlyLeu 353
1423 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1424 AGATCTTCATCTTTACAGAAAGATGC..... 835
1425 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1426 GlnAspValArgCysTyrPhePheIleTyrPheCysGlyCysValLeuAl 370
1427 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1428 AACCTCGAAGACGATGATAT 813
1429 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1430 aArgMetLeuLeuValPheAlaLeuLeuAsnSer.....C 382
1431 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1432 GCACGATTTTGATGAACAGAAATGAATTAACCTTTCTCCGGAATAGT 763
1433 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1434 ySSer..... 383
1435 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1436 TTTTCTTAACAAGTTAATTAATTAATTAATTAATTAATTAATTAAT 713
1437 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1438 CATTAACTTGAACCTTCCAGTGTCCGAGTCAATATCTACACCA 663

```

```

384 ..... ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 .....LeuTrpPheLeuLysGlyPheTyrCysL 393
386 TCTAATCAAAATCATTAACAGCTTCAATATTTATATCC...ATAATC 616
387 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 yAspLeuLeuLeuLysThrLeuMetLeuValSerPheHisPheIleLeu 409
389 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
390 AATTTCTTTTCTAT...ATCATCATATAGGTTTCTCCACCTAAGATA 569
391 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 GluPheLeuPheValCysIlePhePheIlePhePheThrVal..... 423
393 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 AAGAATATTAATTCATCTGAGCTACTTTTCTGATACGTTATTTTCAATTA 519
395 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
396 IleTyrAsnTyr..PheLeuLeuPhePheLeu..... 433
397 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 ACATCATATACATAGTTCATATATCTTAAGCGCTTTCTTAAATAGAA 469
399 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 CysPheValIlePheLysCys 439
401 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
402 TCTTTGCTTCCATTAATGATGATGATACATCATATTAATTAATGTC 419
403 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 spheCysLeu..... 442
405 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 GAGCAATGCAATATATATAAATGACACCATTTGGTTGAATCA..... 375
407 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 AlaAspCysLeuPheLeuLeu.....PheAspPheGlyCysCysLeu 456
409 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 ATCATATGTTTGTCTATACACCTTGACCTTGACCTTGACATATGATATCC 325
411 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
412 IleTyrCysPhe..LeuGlyLeuTyrMetCys.....PheIleL 469
413 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 TGTATTAATACCTTGTTCTCTTTTATTAATCTCTCAAAATTTGTCCTG 275
415 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 euIlePhePheValIleAspPheLeuTyrIlePheValIlePheSerTyr 485
417 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 GTGATTTTCTATCATTTTATTAATTAATTTATTAATTTTGTAAATCTA 225
419 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 CysMetPheTrpSerPheTyrLeuTyrTyrPheAsnPhePheAspIleAl 502
421 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 GAGTTACACCCATCTCCCAATTAAGATCAATATCTACAAATTAATTAAT 175
423 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 aserPheThr..... 505
425 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 ATTCGATTTAAGATGTAAGAAAGACTCATGATTTCCGTTGATGAGTTT 125
427 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 ValPheValIleSerLeuAlaPhe 514
429 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 TATTTTATTTTCCGTAATTAATTAATTAATTAATTAATTAATTTT 75
431 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 IleTyrTyrGlyCysValLeuPhe.....TyrPhePh 525
433 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
434 C.....CTTCAAGGTTCTCGAATTTG 53
435 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 eAsnValAspCysIleMetLeuPheTrpArgIle 536
437 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 seq_name: p1r2:T18482
439 seq_documentation_block:
440 hypothetical protein C0705c - malaria parasite (Plasmodium falciparum)
441 C:Species: Plasmodium falciparum
442 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
443 C:Accession: T18482
444 R:Lawson, D.; Bowman, S.; Barrell, B.
445 submitted to the EMBL Data Library, August 1997
446 A:Reference number: Z18935
447 A:Accession: T18482
448 A:Status: preliminary; translated from GB/EMBL/DBJ
449 A:Molecule type: DNA
450 A:Residues: 1-2437 <LAW>
451 A:Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331931; PIDN:CA811149.1
452 C:Genetics:
453 A:Map position: 3
454 A:Introns: 1397/1

```

A&gt;Note: C0705c

alignment\_scores:  
 Quality: 150.50 Length: 655  
 Ratio: 0.514 Gaps: 38  
 Percent Similarity: 44.733 Percent Identity: 19.847

alignment\_block:  
 US-09-579-383-2 x T18482 ..

Align seg 1/1 to: T18482 from: 1 to: 2437

```

76 AAAAAATATATATATTCATTGGGAATA..ATACGGGAAATATAAA 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1556 LysAsnLysValAspLysIleTyrGluValGluIleArgLysGluLysAs 1572
123 TAAAACTCATCAACGGAAATACATGAGTCTTTTCACATCTTAATGCA 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1572 nLysLysAsnLysAsnArgIleAsnLys..TyrSerTyr.....A 1585
173 ATAAATAGTAATTTGTAGAAATATGATCTTATTCGGAGAT.....GGG 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1585 snAsnLysGluTyrIle.....LeuLysAspLeuArgArg 1596
217 TGTAACTCATCAATTTACAAAAATATATAA..... 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1597 CysGluLysLysIleLysLysAsnAsnLysIleIleIleLysIleAsnAs 1613
247 .....AATATAATATAAAATGATAGAAAATCACCACAGAC 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1613 nLysCysAsnMetLysAsnMetAsnAsnIleTyrSerArgIleLeuArg 1630
281 AATTTTATAGACAGATATAAAAAAGAAACAGCTATTATAGCAGGATAC 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1630 snPheLeuAsnLysAsnLysLysArgLysAsnLysLeuPheSerSerTyr 1646
331 TATGGTTATGGAACAGTCACAGGTGATAGACA.....AACATATGAT 374
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1647 ValAsnIleLeuAsnAspCysLeuHisArgThrIleArgLysHisIleI 1663
375 TGATTCAAACCCAAATGCTGTCATTTTATATATGTCATTTGCTGCATTA 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1663 eLysAsnAsn.....PheIle..... 1668
425 ATATGTTATATGATGTATCTAGACCATTTAATGAGAAACAAATCTCTA 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1669 .....TyrAsnAsnLysGluAsn..... 1674
475 TTAAAGAAACACGGCTTAGAATATGAACCTATGCTATGATGCTTAATGA 524
1674 ..... 1674
525 AATTAGACGTATCAGAAAAGTAGCTCCAGATGATATTCTTTTATGCT 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1675 .....AsnLysIleValGlnAspIleLysGluLeuLysSerL 1687
575 TAGGTGGAAGAACCTATATGATATATAGAA..... 606
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1687 eusSerSerValHisAsnIleArgAspIleGluCysGlyTyrGluGluArg 1703
607 .....AAGAAATTTGATTTATGCTAATAATATATG.....AA 638
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1704 MetArgLysGluMetLeuLysPheAspIleIleIleAspHisLysGluL 1720
639 GCTGTATATGATTTTGTAT.....TAGATGCTAGATATTTGAC. 678
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1720 sLeuIleAsnSerPheAsnPheValCysValAspAsnIleSerLeuGlyL 1737
678 ..... 678
1737 euHisAsnAsnAsnAsnAsnAsnLysAlaLeuGluAsnLysGlyLysSer 1753

```

```

679 .....TCGGAACCAATGGGAGTTTACACCTTAATGAATTAATTT 722
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1754 GluAlaTyrProTyrGlySerAsnLysAsnIleAsn..... 1766
723 TTCAATATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 772
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1767 .....LysAsnIleAsn..IleAsnLysAsnIleAsnI 1777
773 AAGAAAGTTAATTTCAATTTCTGTTTCATCAAAATGCTCATATCAGC 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1777 LeAsnLysAsnIleAsnIleAsnLysAsnIleAsn..IleAsnLysAsnVa 1793
823 GTTTCAGGAGTTGCATCTTCTGTAAAGATG.....AAGATCTCC 863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1793 AsnIleAsnLysAsnValAsnIleAsnGlnAsnIleAsnLysAsnIleA 1810
864 ATATACACTAAATTTTGTCTGAACAATGAACAATGAACAATTAAGATAC 913
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1810 snIleAsnIleAsnIleAsnIleAspLysGlySerAsnMetAsnAsnPro 1826
914 ATAGGCGACAGCAGATGTTATCAGCAGAACTTTT..... 948
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1827 CysAsnHisGlnAsnLysAsnArgGluAspIleHisTyrAsnAsnHisTy 1843
949 .....ATTAATATTTTATACAGCAAGAGGAAATATAGATCTTG 988
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1843 rLysHisMetIleTyrPheHisAsnLysHisAsnAspLysSerGluLeuL 1860
989 TA..... 990
1860 euLysLysAsnSerGlyGlyLeuGluTyrAsnHisGluGluArgLysTyr 1876
990 ..... 990
1877 LysGluLysIleTyrSerCysLysArgAsnPheGluIleGluAsnIleAs 1893
991 .....TTTATTCAA.....ACATACATTTAGAAA 1015
1893 nTyrPheIleTyrArgAsnIleTyrArgProIleLysTyrHisTyrAspI 1910
1016 CTACAATCCAGATATATAGTATGATGATCTTATCCCAT..... 1056
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1910 IeThrAsn..LysIleLeuIleAsnSerPheLeuIleAsnMetAsnLeu 1925
1057 ..TTATATTTTGGTTTAAATATATACATCAACATCATATTAGTTTTC 1103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1926 LeuIleTyrPhe.....TyrLeuLeuLysCysIleLeuPheAspAs 1939
1104 ATTAGAACATACAGAGGTGGA.....TTTAGTCCCGAAA 1138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1139 ATTAAGAATTTATAGAAATTTGTAGAAAACAAATATACATGATATAAAC 1188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1956 snLys.....IleIleAsnLys..IleGluAsnAspAspTyr 1967
1189 AATAATAATAGGCGAGATGTTATAGGATATGCAATTTATTTATGAAGA 1238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1968 AsnAspAsnAsnLysAsp.....HisIlePheValAsnLys 1979
1239 ACAATTTACCAACTGATCATTCGATGTAGATATTTTCTTACA...ATA 1285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1979 s.....AsnLysHisMetValGlnTyrPheLeuAsnLysAsn 1992
1286 TTTCGAAACATTTAAATCCCTGAGTACAAACTCCAAAGACCTTACTATA 1335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1992 heTyrThrSerCysAsn.....IleLysAsnLysCysVal 2003
1336 ACTGAAACCCCT.....GAAGACTGTAGCAACATAGATGATATGCT 1376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2004 AspLysAsnLysLeuTyrTyrThrAspIleSerThrIle..... 2016
1377 TCCAGGACTCGTTATTCACACCATAGGATATATTAACAACAAATGATG 1426

```







```

154 yspIovallleasnasnlyrilleasnlguleuasnillelyrlyrille 170
156 .....
171 leuSerCysTyrglySerlysllylleaspthAlaleuLysh 187
196 .....
1965 .....GTAATTAATAATTAATAAGTCTGCTGATA 932
187 fAlaspserglyTyrlleuthrlysrgrleuileasnillethnSerAsnp 204
931 ACATGCGTCGCCCTA...TGAATCTTTATTTGTTCTATTTGCTCA 885
204 heillellelyslleuasnlySerProphelle...leuLystry 219
884 GACAAAATTTAGTGTATATGAGAT..... 858
220 lleuasnmetaspilletyrlysllysnilleleuProleuasnille 236
857 ....TCTTCATCTTTACAGAAAGATGCACTCTGAAACGATGATATG 812
236 uArgphelysilleuGlnAsnAsnille..... 245
811 CAGCATTTGATGACACAGAAATTTGAATTTCTTCCGGAATA... 765
246 .....leuasnleuasnlythrphe 253
764 .....GTTTCTTAACAGATTAAT 745
254 lleTyrllyrlysnthlyrlylethlyrlyrlyleleuasnlyseu 270
744 TAAATTAATTAAT.....AATTTGAAA..... 722
270 uasnleu...tyrAsnArgrArgrAsnilleTyrlleuasnilleLysserVallyr 286
721 .....AATTAATTCATTTAAGTGTAAACT..... 695
287 leuCyssasnilleTyrlAsnAsnilleCyssasnlyrlysl 303
694 .....TCCCATGTGCT.....TCCCATG 677
303 nleuTyrllystryrAsnleuGlyGlnhislleglyVallelserSerGln 320
676 CAATATGACACACATCTAATCAAAATCTTAACAAGCTCAATATTTTA 627
320 laileSerGlnProserThrGlnMetValleuArgrThrphelhis..... 334
626 TCCACATATCAATTTCTTTCTATATCTATCATATAGATTTCACACC 577
335 .....AlaserSerlleu..... 339
576 TAAGGATAAAGAAATTAATTAATCTGAGCTATTTTCATAGCTGTAA 527
340 .LyAspLysspHeasnPhasnlyrlyleuileTyrllysllyrlyleu 356
526 TTTTATTAAGATCATACCA.....TAGTTTCATATTTCTAG 489
356 yrllyslleuasnilleasnlysllephelysleuilelleasnphelysly 372
488 CCGGTTTCTTAATAGGATCTTGTCTTCATTAATGCTGTATGATAC 439
373 TyrlleasnillelyspHeasnlylephleuMetVallysllyleuTy 389
438 ATCATAT...AACATATTAATGCGAG.....CAAAATGCAA 407
389 rAsnTyrlAsnAsnilleleuPhe.GluTyrllystryrilleuGlnsnGln 405
406 TATATAAATTAATGACACCATTTGGGTTGAATCATATATGTTTCTCTA 357
406 TyrllelysllysnphelilleTyrlAsnserlleserlysnphelysly 422
356 TCACCTTGACCTGTCATGAAACATAGATCTGCTATTAATACCTT... 311

```

```

422 rAsnleu.....AsnAsnillellellelystryrleuAsna 434
310 .....GTTTCCTTTT...TATACCTCTCA..... 287
434 snValillelystryrlyrAsnTyrlSerAsnilleGlnleuileuileysn 450
286 .....AAATTCCTTGGTATTTTCATATTTATTTATTTATTT 246
451 lleHlAsnlystryrilleuTyrlAsnilleTyrllyrlyrlyrly 467
245 TTAATATTTTTCATATCTAGATTACACC..... 215
467 rtyrlysllyspHeTyrlAsnleuTyrlAsnlysllylleleuAsna 484
214 .....CATCTCCGCAATGATCCAT 194
484 snAsnAsnAsnlystryrAsnValilleTyrlphleuileasnlyrphesn 500
193 ATTCACAAATTAATTAATTTTCATTTAGAT.....GT 159
501 leuPheserAsnlyrlyrlyrlysllyrlyrlyrlyrlyrlyrlyrly 517
158 GAAAAAGACTCAGATTTTCGTTGATGAGATTTATTTTA.....TT 115
517 easnSerAsnlyr...TyrlphelyslyMetAsnphelle...leuLyssAsnp 533
114 TTCCTGATTT...ATTCCAATGAATTAATTAATTA.....TTTTTC 74
533 easnAsnilleGlnilleuasnlyslleuPhelyrlyValAsnAsnillephel 550
73 CTTCAGAGTTCTGCAATTTGCAGAAATCAAGATGATCTTAATTTAA 24
550 leTyrllystryrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrlyrly 566
23 AATATGATATTTTAAATTC 3
567 Asnillellellelysllystryr 573

```

seq\_name: plr2:T43916

seq\_documentation\_block:  
 chitinase A [imported] - Pyrococcus kodakaraensis  
 C:Species: Pyrococcus kodakaraensis  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43916  
 R:Tanaka, T.; Fujiwara, S.; Nishikori, S.; Fukui, T.; Takagi, M.; Imanaka, T.  
 Appl. Environ. Microbiol. 65, 5338-5344, 1999  
 A:Title: A unique chitinase with dual active sites and triple substrate binding sites  
 A:Reference number: 222722; MUID:20049967  
 A:Accession: T43916  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1215 <TAN>  
 A:Cross-references: EMBL:AB024740; NID:g6580044; PIDN:BAA8380.1; PID:g6580045  
 A:Genetics:  
 A:Gene: chIA

alignment\_scores:  
 Quality: 148.00 Length: 647  
 Ratio: 0.576 Gaps: 28  
 Percent Similarity: 39.722 Percent Identity: 17.311

alignment\_block:  
 US-09-579-383-2 x T43916 ..

Align seg 1/1 to: T43916 from: 1 to: 1215

```

160 CATCTTAATCAATTAATAGTAATTTTGAGAAATGATGATCTTATGCGG 209
139 HisleuLyssAlaAsnThrThyTyrllyrlyrlyrlyrlyrlyrlyrlyrly 155
210 ACATGGGTCTAATCTAGCAATTAACAAAAATTAATAATTAATAA 259

```



1580 AAGGGGAACATATTAAATGCGCACAAGGACCT 1620  
 |||:|||||  
 683 hrProValSer.....TrpAsnArgGlyPro 691  
 seq\_name: p1r2:S60651

seq\_documentation\_block:  
 chitinase precursor - Serratia marcescens  
 C:Species: Serratia marcescens  
 C:Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 10-Dec-1999  
 C:Accession: S60651  
 R:Bruberg, M.B., Eijsink, V.G.H., Nes, I.F.  
 FEBS Microbiol. Lett. 124, 399-404, 1994  
 A:Title: Characterization of a chitinase gene (chiA) from Serratia marcescens BJL200 and  
 A:Reference number: S60651; MUID:95154677  
 A:Accession: S60651  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-563 <BRU>  
 A:Cross-References: EMBL:236294; NID:908884; PIDN:CAA85291.1; PID:908885  
 C:Superfamily: Serratia marcescens chitinase

alignment\_scores:  
 Quality: 145.50 Length: 425  
 Ratio: 0.731 Gaps: 20  
 Percent Similarity: 46.824 Percent Identity: 20.471

alignment\_block:  
 US-09-579-383-2 x S60651 ..

Align seg 1/1 to: S60651 from: 1 to: 563

```

25 TTAATTAGTATCCATCTGTAATTCGAATTCAGACCTGAAAG 74
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 LeuLeuIleGlySerThrLeuCySerAlaIleGlnAlaIleAlaProG 27
75 AAAAAAATAATA.....AATAATTCATTGGGAATAATACGGGAA 115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 YLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44
116 ATAAATAATTAACATCAACGCAATACAGTCTTTTTCACATCTT 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 spGlnAlaIleThr.....AlaTrpAsnSerLeuValLysVal 56
166 AATCGATAATAGTAATTTGTAGAAATAGATCTTATTCGGGAGAT 213
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 LysAsnAlaIleAspValSerValSerTrpAsnLeuTrpAsnGlyAspTh 73
214 .GGGTGTAAGTCTAGAAATTCACAAAATAATATAA..... 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 rGlyThrThrAlaLysValLeuLeuAsnGlyLysGlnAlaTrpSerGlyP 90
247 .....AATATAATTAATAATGAT 264
90 roSerThrGlySerSerGlyThrAlaAsnPhelYsValAsnLysGlyGly 106
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 AGA..... 267
107 ArgTyrGlnMetGlnValAlaLeuCyAsnAlaAspGlyCysSerAlaSe 123
267 ..... 267
123 rAspAlaThrGluIleValAlaIleAspThrAspGlySerHisLeuAlaP 140
268 .....AATACCAAGACAAATTTTACAGAGCTATAAAAAAGAAACAA 312
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 roLeuYsGluPro.....LeuLeuGlnYsAsnLysProTyrLysGln 154
313 GGT.....ATTATAGCAGGATAGTACTGATGCTCATGCAACAGTCAAG 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 AsnSerGlyLysValValGlySerTyrPheValGluTrpGlyValTyrGln 171

```

```

354 TGATAGCAAAAAACATATGATTCATCAAAACCAATGTCATATTTAT 403
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 Y.....ArgAsnPhetThrValAspLysIleProAlaGlnAsnLeuThr 186
404 ATATTGCATTTCCTGCATT.....AATATGTTATATAT 438
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 ILeuLeuTyrGlnPheIleProIleCyGlyLysGlnLysAlaAsn 202
439 .....GATCTGAGCACTTTAA 455
203 SerLeuLysGluIleGluGlySerPheGlnAlaLeuGlnArgSerCysG 219
456 TCGAAGACAAAGATTCTTATAGAAACAC.....GGCTTAC 493
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 nGlyTrpGlnAspPheLysValSerIleHisAspPheAlaIleLeuG 236
494 AATATGAAACCTATGTATGATG..... 516
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 ILeuYsAlaGlnLysGlyValThrAlaTrpAspAspProTyrLysGlyAsn 252
517 CTATAGAAATAGACGTATCGAAAGATCGATGTAATTAATTCCT 566
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269
567 TTTATCCTTAGCTGGAAGAAC.....TATATGATAG 598
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 uProSerIleGlyGlyTrpThrLeuSerAspProPhePheMetGlyA 286
599 ATATAGAAAAACAAATTCATATGTCGATTAATATTAACCTGTATAT 648
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 sPlyValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302
649 GATTTTGAT...TTAGATGTCGTAGATTCATGACGCGGAA...CCACATGG 692
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyG 319
319 YLysGlyLysAlaAsnProAsnLeuGlySerProGlnAspGlyGlnThrTyrV 336
737 TTAATTTATTAATCTGTTAGAAAAACTATTCGGAAGAAAGTAAAT 786
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 aLLeuLeuMetLysGluLeuArgAlaMetLeuAspGln..... 348
787 TCAATTTCTCGTTTCATCAAAATCGTCATTCATGCGGTCGAGAGTTGC 836
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 .....LeuSerAlaGlnThrGly..... 354
837 ATCTTTCTGTAAGATGAGAATCTCATATATACACTAAATTT...TTGT 883
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 .....ArgLysTyrGlnLeuT 360
884 CTGAACAAATAGAACAAATTAAGATTAATACATAGGACGACGAGATGTA 933
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374
934 TCAGCAGCACTTTTATTAATATTTTAATACAGCAAGGAGAAATAGA 983
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 .....TyrAsnValAlaGlnAsnSerMetAs 383
984 TCTGTATTTATTCAAACATACAAAT 1008
|||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 PHisIlePheLeuMetSerTyrAsp 391

```

seq\_name: p1r2:A38221  
 seq\_documentation\_block:  
 chitinase (EC 3.2.1.14) MFI - nematode (Brugia malayi)  
 C:Species: Brugia malayi  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
 C:Accession: A38221  
 R:Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Plessens, W.F.; Perler, F.B.



A:Cross-references: GB:AE001403; GB:AE001362; NID:g3845216; PIDN:AAC71902.1; PID:g3845216  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PFB0565w

alignment\_scores:  
 Quality: 144.00 Length: 558  
 Ratio: 0.514 Gaps: 31  
 Percent Similarity: 50.179 Percent Identity: 20.968

alignment\_block:  
 US-09-579-383-2 x H71611 ..

Align seg 1/1 to: H71611 from: 1 to: 1817

```

64 ACCTTGAAGAAAAAATAATATATATCATTTGGTAATAATACGGA 113
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
355 ThrLysGlnGlyLysHisAspMetAsnAsp.....TY 365

114 AAATTAATAAATACATCATCAACGAAATACAGCTTTTTCACATC 163
   ||| ||||| :|||:|||||:|||||:|||||:|||||:|||||:
365 rAsnMetAsnLysLysAsnAsnMetAspIleAsnIleThrIleAsnAsnA 382

164 TTAATCGAATATAGTAATTTGTAGAAATAGATCTTATTTGGGAGAT 213
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
382 snAsnAsnAsnAsnAsnAsnLysIle.....TyrAsnAspAsn 394

214 GGGTGAACCTCTAGAAATTCACAAAATAATATAATATAATATAA.... 258
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
395 ThrLeuAsn.....ValTyrAsnAsnSerTyrAsnIleHisSerAsnH 409

259 .....AATGATGAATAATCACCACAGCAATTTTAGAGAGATATAAA 301
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGlnLysHisLeuL 426

302 AAAGGAACAAGCTATTATAGCAGACTACTATGCTCATGGAACAGTCAA 351
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
426 ys.....MetLeuCyAspAsnPhe.....PheAsnLeuGln 436

352 GGTGATAGACAAACATATGATGATTCACACCAATGGTGTCAATTTT 401
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
437 GluPheTyrSerSerAsnIleIleIleAsnAsnMetAspIleGlnTyrTh 453

402 ATATATGATGCTTCGCATTAATATGATGATGATGATCTAGACCAT 451
   ||| ||||| :|||:|||||:|||||:|||||:|||||:|||||:
453 rTyrAspTyrPhe.....IleLeuTyrGln.....LysCysP 464

452 TTAATGGAAGACAAGATTCCTATTAAAGAAACACGGCTTAGAATATGAA 501
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
464 heLeuProIleGlnArgIleVal.....HisValAsnTyrMet 476

502 ACCATGATGATGATGCTTAAATAGACAGTATCAGAAAGATGATGCC 551
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
477 LysTyr...LeuTyrLysAsnAsnGlnArgLysLysAsnLysIleArgLys 492

552 AGATGTAATATTCTTTTATCCTTAGGTGGAGAAACCTATATATGATGA 601
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492 sPheLeuIleThrLeuLeuGln.....TyrSerArgAspI 504

602 TAGAAAAAGAAATGATATGATGATGAAATATG...AAGCTTTGTAAT 648
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
504 leAsnPheHisPhePheIlePheAsnLeuIleLeuTyrLysCysLysAsn 520

649 GATTTTGATTTGATGCTGATAGATATGCTGGGAAACACATGGAAGTT 698
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 GluPheProCysSerIlePheGlnLeuHisIleSerGlnTyrLeuTyrPh 537

699 TTAC...AAGTTAATGAATTAATTTTCAATATATATAT..... 738
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
537 ePheValLysLeuAsnGlnLeuAsnIleLysAspAlaTyrIleTyrTyrP 554

739 .....AATTAATTAACCTGTTAAGAAAAACT 765
```

```

554 heAsnAsnPheLysTyrGlnAspMetIleIleTyrPheSerArgLysAla 570
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 ATT...CCGGAAGAAAAGTAAATTTCATTTCTGGTTCATCAATGCTGC 812
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
571 PheTyrProTyrPglu.....ThrAsnValGlnGlnGlnLysGlnTh 585

813 ATTTATCATGCGTTTCAGAGATTGCATCTTCTGTAAGATGAAGAAATCTC 862
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 rLeuSerTyrIle..... 589

863 CATATAACACTAAATTTTGTCTGACAAATAGAAACAATAAAGAAATTA 912
   ||||| ||||| :|||:|||||:|||||:|||||:|||||:|||||:
590 ..TyrAsnAspLys.....IleLysLysAsnLysLys... 599

913 CATAGGGCAGCAGGATGTATTCAGACAGAACTTTTATAT..... 954
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 AsnAsnSerTyrTyrGlnMetAsnAsnAsnThrTyrMetAsnGlnHisG 616

955 .....ATTTTAATATACAGCAAGAGAGAAATAG 982
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
616 yTyrThrAspIleGlnAsnGlnArgLeuAsnLysLysAsnLysArgLeuA 633

983 ATCTTGATTTTATTCACAACATACATTTAGAACTACAAATCCAGATATA 1032
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 snVal.....ArgLysArgTyrAsnThrLeuAspAspIle 644

1033 ATGCTAGATATGCTATTCACCATTTATTTTGGTTAAATATATACAT 1082
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 IleValSer.....AspHisGlnLysAsnSerTyrAspLysTyrAsnTh 658

1083 CACAAATCATATAGCTTTTCATTAGAACATTAACAGAGGTGATTAAGTC 1132
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
658 rSer.....LysHisAsnArg..... 663

1133 CCGAAATTAAGAATTTTATGAAATGGTAGAAACAATACATGATATA 1182
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 ....ArgLysAsnHisIleAsnGlnMetLysLysLysGlnAsnAsnLys 678

1183 AATCAAAATATATATAGGCGAGATGATAGATATGCGATTTGATTTAT 1232
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
679 LysLysAsnThrLeuPheValAspGlnLysAspMetGlnGlyIleGlnLys 695

1233 GAAAGAACAAATTCACCACTGATCATTCGATGATGATATTTTCTTACA 1282
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
695 sGlnLysGlnLysGlnAsnLysAsnMetAsnAsnIlePheTyrAsnA 712

1283 AT.....ATTTGGAACAATTTAAATCTGTGA 1308
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
712 snSerTyrSerAsnIleAsnAsnSerSerTyrSerAsnIleAsnAsnAsp 728

1309 GTACAAACTCCAAAGACCTTACTATACCTGAAGAAACCTGAGACTGTAG 1358
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
729 IleTyrSerValAspAsnMetThrSerValAsnAsnThrLysTyrValSe 745

1359 CACAATAGATGATATGCTCCAGAGACTCGTTATTCACCAATAGGAGAT 1408
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 rGlnValProSerTyrAlaHisValIleLeuAsnLysGlnValAsnGln 762

1409 ATTACAAA.....CACAAATGATGCTATATGGAAGAACTGATGCT 1446
   ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
762 yTyrGlnGlnLysLeuProAsnTyrAsnAsnMetMetLys...GlySer 777

1447 TATTCATTTATGACACCTGCTGTAGACAGATATGATGAGACTGTGCAA 1496
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778 HisIleIleAsnGlnLeuProLysAsnAsnTyr..... 788

1497 AGTATGCTATGAAAAAATATGCGATGGGAAGACCCATTAAT..... 1539
   ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
789 ....IleTyrGlnAsnAsnTyrIleGln.....AsnTyrLeuMet 802

1540 .....TATAACACTGACTATTAAGAAAGCTATATTTATATG 1578
   ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
```







```

392 snSerCysLeuTrpPheLeuGlyPheTrpCysLysAspLeuLeu 408
651 ATC...ATTACAAGCTTCATATTATTCACATATTCATTC... 610
      ::::: ||| |||: |||: |||:
409 LeuGlySerLeuMetLeuThrSerPhePheLeuLeuGlyLeuLeu 425
609 .....TTTTCTATATCTATCATATAGTTTCTCCACCTAAGATAAA 567
      ::::: |||: |||: |||: |||:
425 sValCysLeuPhePheLeuPhePhe..... 433
566 AGAATATATACATCTGACGACTTTCTGATACGCTATATTCATTAG 517
      ::::: |||: |||: |||: |||:
434 .....ThrValIleTrpAsnTrpPheLeuLeuPheLeu..... 445
516 CATCATACCATAGTTTCATATCTTAAGCGTGTTCCTTAATAGCATC 467
      |||: |||: |||: |||:
446 .....CysPheValPheLysCysP 452
466 TTGTCTTCCTTAATAGTCTAGATACATCATATTAATATGCGA 417
      ::::: |||: |||: |||: |||:
452 heCysLeu.....ValAspThrLeuPheLeu..... 461
416 GCAATATGCAATATATAATGACACCATGGTTGATCATCATATG 367
      |||: |||: |||: |||:
462 .....PheAspPheGlyCysC 467
366 TTGTGCTCTACCTGACTGACTGTCATGACCATAGTATCTGCTATAA 317
      |||: |||: |||: |||:
467 S.....LeuV 469
316 TACCTGCTTCCTTTTATCTCTCTAATAATTTGCTGTGATTTT 267
      ::::: |||: |||: |||: |||:
469 alTrpCysThrPheCysLeuTrpMetCysPheValLeuLeuPhePheVal 485
266 CTATCATTTTATTTATTTT 243
      |||: |||: |||: |||: |||:
486 LeuAspPheLeuTrpValPheIle 493

```

seq\_name: p1r2:A25090

seq\_documentation\_block:  
 chitinase (EC 3.2.1.14) - *Serratia marcescens*  
 C:Species: *Serratia marcescens*  
 C:Date: 12-Feb-1988 #sequence\_revision 12-Feb-1988 #text\_change 10-Dec-1999  
 C:Accession: A25090  
 R:Jones, J.D.G.; Grady, K.L.; Suslow, T.V.; Bedbrook, J.R.  
 A:Title: Isolation and characterization of genes encoding two chitinase enzymes from *Serratia marcescens*  
 A:Reference number: A25090  
 A:Accession: A25090  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-561 <ON>  
 C:Superfamily: *Serratia marcescens* chitinase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

alignment\_scores:  
 Quality: 142.50 Length: 425  
 Ratio: 0.720 Gaps: 20  
 Percent Similarity: 46.588 Percent Identity: 20.471

alignment\_block:  
 US-09-579-383-2 x A25090 ..

Align seg 1/1 to: A25090 from: 1 to: 561

```

25 TTAATATATATCATCTGTATCTGCAAAATTCAGAACCTTGAAG 74
|||: ||| ||| |||: |||: |||: |||:
11 LeuLeuIleGlySerThrLeuGlySerAlaIleGlnAlaIleAlaProG 27
75 AAAAAATATATA.....AATATTCATTTGGAAATATACGGCAA 115
||| |||: |||: |||: |||: |||: |||:

```

```

27 YLSPProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44
116 ATAAAAATAAACTCATCAAAACGGAATACATGAGCTTTTTCACATCT 165
      ::::: ||| |||: |||: |||: |||:
44 spGlnAlaIleThr.....AlaTrpAsnAsnLeuValLysVal 56
166 AATCGAATATAGTAATTTGTAGAAATATGATCTTATTTGCGAGAT 213
      ::::: |||: |||: |||: |||: |||:
57 LysAsnAlaIleAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73
214 ..GGGTAACTCTAGATTTACAAAAAATATATAA..... 246
      |||: |||: |||: |||: |||: |||:
73 agLYThrGlyProLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90
247 .....AATATAATATAAATATAT 264
      ::::: |||: |||: |||: |||:
90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106
265 AGA..... 267
107 ArgTrpGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAla 123
267 ..... 267
123 rAspAlaThrGluIleValValAlaAspThrAspGlySerHisLeuPro 140
268 .....AATCAACCAAGCAAAATTTTAGAGAGATATAAAAAAGCAACAA 312
      |||: |||: |||: |||: |||: |||:
140 roLeuLysGluPro.....LeuLeuGluLysAsnLysProTrpLysGln 154
313 GGT.....ATTATAGCAGATACTATGCTTCAAGACACTCAAG 353
      ::::: |||: |||: |||: |||: |||:
155 AsnSerGlyLysValValGlySerTrpPheValGluTrpGlyValTrpG 171
354 TGATAGAGCAAAACATATGATGATTCACCAACCGATGGTGCATTTAT 403
      ::::: |||: |||: |||: |||: |||:
171 Y.....ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThr 186
404 ATATTGCATTTGCTCGCAT.....AATATCTTATATGAT 438
      ::::: |||: |||: |||: |||: |||:
186 ILeuLeuTrpGlyPheLeuProIleCysGlyGlyAsnGlyIleAsnAsp 202
439 .....GTATCTACACCATTTAA 455
      ::::: |||: |||: |||: |||: |||:
203 SerLeuLysGluIleGlySerPheGlnAlaLeuGlnArgSerCysG 219
456 TCGAAGCAAAAGATCTCTATTAAGAAAAC.....GGCTAG 493
      ::::: |||: |||: |||: |||: |||:
219 ngLYArgGluAspPheLysIleSerIleHisAspProPheAlaIleLeuG 236
494 AATATGAACCTATGATGATG..... 516
      ::::: |||: |||: |||: |||: |||:
236 IuLysAlaGlnLysGlyValThrAlaTrpAspAspProTrpLysGlyAsn 252
517 CTTAATGAATATGACGTATCAGAAAAGTACGTCAGATGTAATATTTCT 566
      ::::: |||: |||: |||: |||: |||: |||:
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269
567 TTATCTCTTAGCTGAGAAC.....TATATGATAG 598
      |||: |||: |||: |||: |||: |||:
269 uProSerIleGlyTrpThrLeuSerAspProPhePheMetGlyA 286
599 ATATAGAAAAAAGAAATGATATGATGATAAATATTAAGACTTGTAAAT 648
      |||: |||: |||: |||: |||: |||: |||:
286 spLysValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302
649 GATTTGAT...TTAGATGCTGATATGATGCTGCA...CCACATGC 692
      ::::: |||: |||: |||: |||: |||: |||: |||:
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyG 319
693 GAAGTTTAC.....AAGTTAATGAATTAATTTTCAATATATATA 736
      |||: |||: |||: |||: |||: |||: |||:
319 yLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluThrTrpV 336

```

```

737 TTAATTAATTACTGTTAAGAAAACATATCCGAGAAAAGTTAATT 786
   :: |||::: |||:::
336 alleuMeuMetLysGluLeuArgAlaMetLeuAspGln..... 348
787 TCAATTTCTGGTTCATCAATGCGCATTTATCATCGCTTCAGAGAGTTGC 836
   |||::: |||:::
349 .....LeuSerAlaGluThrGly..... 354
837 ATCTTCTGTAAAGATGAGAAATCTCCATTAATTAACCTAAATTT...TTGT 883
   |||::: |||:::
355 .....ArgLysTyrGluLeuT 360
884 CTGACAAATAGAACAAATAAAGATTTACATAGGCGCAGCGCATGTTA 933
   ::|||::: |||:::
360 hsrerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374
934 TCAGCAGAACTTTTATTAATTTTAAATACAGCAAGAGAAAATAGA 983
   ::|||::: |||:::
375 .....TyrAsnValAlaGlnAsnSerMetAs 383
984 TCTGTATTTATTCAAACATACAAAT 1008
   |||::: |||:::
383 pHisIlePheLeuMetSerTyrAsp 391

```

seq\_name: p1r2:B71621

```

seq_documentation_block:
  probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
  C/Species: Plasmodium falciparum
  C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
  C/Accession: B71621
  R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
  Science 282, 1126-1132, 1998
  A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
  A/Reference number: A71600; MUID:99021743
  A/Accession: B71621
  A/Status: preliminary; nucleic acid sequence not shown; translation not shown
  A/Molecule type: DNA
  A/Residues: 12295 <GAR>
  A/Cross-references: GB:AE001379; GB:AE001362; NID:g3845118; PIDN:AAC71827.1; PID:g384511
  C/Genetics:
  A/Experimental source: clone 3D7
  A/Gene: PFB0190c

```

```

alignment_scores:
  quality: 139.50      length: 731
  ratio: 0.431         gaps: 37
  percent similarity: 44.323  percent identity: 18.741

```

alignment\_block:  
US-09-579-383-2 x B71621 ..

Align seg 1/1 to: B71621 from: 1 to: 2295

```

1 ATGAATTTAAAAATATCAATATTTTAAATATAGTATCCATC..... 42
   ::|||::: |||::: |||:::
8 LeuAsnPheIlePheSerPheLeuLeuIleIleLysCysAspGluSe 24
43 .....TTGTATCTGCAAAATTCAGAACCTTGA 70
   ::|||::: |||:::
24 rValSerAsnGlyArgLysGluIleTyrPheAspAspGluLysLeuL 41
71 AA.....GGAAAAATATATTAATTAATTCATTCATGGCA 102
   |||::: |||::: |||:::
41 yLseuSerSerPheAspArgSerThrAsnIleAsnLeuAspValGly 57
103 .....ATAATACGGGAA.....ATAAAAA 122
58 GluAsnAspGluLeuSerSerTyrValProArgGluValAspGluLys 74

```

```

123 TAAACATCATCAAGCAAAATACATGAGCTTTTTCACATCTTAATGCA 172
   |||::: |||::: |||:::
74 sLysAsnLysLysAspIle...AspSerLysGluAsnSerLysSerG 90
173 ATAAATAGTAATTTTGTACAAATATGATGATCTTATTCGGAGATGGTGAAC 222
   |||::: |||:::
90 LysAsnAsnIleTyr.....Asn 95
223 TCTGAATATCAAAAAATATATAATATAATATAAT..... 261
   ::|||::: |||::: |||:::
96 LysAspAsnThrLysAsnAsnGluAspValAsnTyrAsnValIleuLy 112
262 .GATGAAAATATCCACACACAAATTTTGAAGAGAGTATAAAAAGAAC 310
   |||::: |||::: |||:::
112 sAspGlyArgAlaLysGluGlyIleIleThrAsp...GluLysArgArg 128
311 AAGGATATATAGCAGAGATACATAGTTCATGAGAACAGTCAAGGTGATGA 360
   ::|||::: |||::: |||:::
128 erSerThrLysAspGlyLysAsnLysGluGlnAsnAsnLysMetAsn 144
361 GCAAAACATATGATGATTCATCAAAACCAATGCTCATTTTATATATTGC 410
   ::|||::: |||::: |||:::
145 SerAspAspValHisAspAsnAsn..... 152
411 ATTGCTCGCATTTAATATGTTATATGATGTATCTAGACCATTAATGAA 460
   |||::: |||::: |||:::
153 .....AsnAsnMetAsnAspIleAsn..... 159
461 GACAAAGATTCCTATTAAGAAACACGCGTTAGAAATGAAACCTATGCT 510
   |||::: |||::: |||:::
160 .....PheValValGlu.....TyrAsn 165
511 ATGATGCTTAATGAATATAGCATATCAGAAAGTACGTCAGATTAAT 560
   |||::: |||::: |||:::
166 LysMetIleAspAsnTyrAspLysIleLeu.....AspGluLe 178
561 TATTCCTTTATCCCTAGGTGAGAAACCTAT.....ATGATAG 598
   |||::: |||::: |||:::
178 uIleLeuLysSerIleAsnArgAsnAsnTyrAsnTyrPheAsnMetLeu 195
599 AT.....ATGAAAAAGAAATTT...GATTAATGTC 624
   |||::: |||::: |||:::
195 spGluTyrSerLeuGlnThrLysLeuAsnLysGluMetLysAspSerLe 211
625 GATTAATAATATGAAGCTTGTAAATGATTTGATTTAGATGCTGATAGAT 674
   ::|||::: |||::: |||:::
212 AsnTyrLeuIleArgLysMetAsnLysAsnSerArgLysTyrPheI 228
675 TGACTGGGACACATGGGAGTTTACAACTTA...AATGAATTAAT 721
   |||::: |||::: |||:::
228 eSerPheSerAsnAsnGluLysLysLysIleIleLysAsnAspMetAsn 244
722 TTTCAAATTAATTAATTAATAA.....TTAATTAATCTGTTAAGAAAAC 765
   ::|||::: |||::: |||:::
245 ..GluAsnIleTyrIleArgHisPheIleValSerLeuThrAsp... 259
766 ATTCGGAGACAAAGTTAATTTCAATTTCTGCTCATCAATGCTGCAT 815
   ::|||::: |||::: |||:::
260 TyrAsnAsnPheLysLeuIleGluThrCysPheAspLysAsnAsnPhel 276
816 ATCATCGCTTTCAGAGGTGTCATCTTCTGTAAAGATGAAGATTCAT 865
   ::|||::: |||::: |||:::
276 eTyrTyrIle.....AspGluAsnLysIleT 285
866 ATTAACACTAAATTTTGTCTGAACAAATAGAAACAAATTAAGATTAAT 915
   |||::: |||::: |||:::
285 ySerTyrLysTyr.....AsnTyrLysLeu... 293
916 AGGCGACAGCAGATTTATCAGCAGAACTTTATTAATTAATTTTAATAC 965
   ::|||::: |||::: |||:::
294 ..MetLeuAsnLeuPheSerSerGluAsnPhelLeuTyrTyrIleAsn 309
966 AGCAAAAGCAAAATATGATCTGTATTTATTCAAACATACAAAT..... 1008

```

```

309 userlysheserleu...leugluileleaspsasnlyrasnlystyrs 325
1009 .....TTAGAACTACAAATCCAGATATATATAGTATATAGTACTTA 1050
325 erpheiileleasnsnleleystargasptryproasnsnmetlytyal 341
1051 TCCATTATATATTTGGTTTAAATATATACATCACATCATATATAGTTT 1100
342 CysglInseryphenetyraspheiiletyrser .....Ty 352
1101 TTGATTAAACATATACAGAGGTGGATTAGTCCCAAAATAAGATTA. 1149
352 rphenleuserlyrasnhs...Hisphenaspshslyshslystyryleui 368
1150 .....TTAGAAATGGTAGAAAAACATACATGATATATATATATATAT 1194
368 leasmetaspilietrasnsnserileglnthlysglyln..... 382
1195 AATAGGCGAGATGTATATAGGATATGCGATTTATTTATGAAA..... 1236
383 .....ileglyashhslyseuTyrlslylslyleuLysly 394
1237 .....G 1237
394 sleuasngluasnleuileuTyrasnTyrllelysasnspsSerg 411
1238 AACATTACCACTGATCATTCGATGATATGATTTTCTTACAAATATTT 1287
411 lmetileprotyvalthrleuGlumetargmellipheserasnph 427
1288 TGGAAA..... 1293
428 ThrasnleuileaspilleuasnlyseuTyrasnleasptyrGl 444
1294 .....CATTAAATCCCTGAAGTACAACTC 1318
444 naspsnlelelysglngluasnvalasnvalasnProglmragspalap 461
1319 CAAAAGACCTTACTATATAGTAAAAACCTCGAGACTGTAGCAATAGAT 1368
461 roglmrapslyrValHisasnlysasnspsvalaspsalSerleuLysasn 477
1369 GAATATGTTCCAGACACTGTTATTCCAACCATAGGATATATTACAAACA 1418
478 VallysgluProlysglyVal.....GluH 486
1419 CAATGATGCTATATGAGAACTAGATCTTATTCATTCATGACACTGGTG 1468
486 sasnlysalmet.....SerAsnTyrgluThrapsgluaraglyA 500
1469 TAGACAGATATGATGGAGCTTGGTCAAGATATGCTATGAAAAATA... 1515
500 spmetiletyrAsp...AsnthrasnlysgluLysPhegluLysserGlu 515
1515 ..... 1515
516 GlyThrphenasnleleSerglygluLyspThrphenlysasnleleS 532
1515 ..... 1515
532 rGlygluLyspThrphenlysasnleleSerglygluLyspGlyluVal 549
1516 .....TCGATGGGAAAGCAGCC..... 1533
549 spGlyAspGlygluLyspGlyAspGlyAspGlygluGlyAlaspsasp 565
1534 .....CATTATTATACACTGACTATTAAGAAAGCTCTAT 1568
566 SerSerValasprThrHisasnlysasnsnlysglyLysgluSergluse 582
1569 TATTTATGAG..... 1578

```

```

582 rAspValThrpsnleuLeuMetaspSerTyrlslylsleuAlaasnspg 599
1579 .....AAAGGGAAACCATATTTAATTAATGCTGGCAACACAGCA 1617
599 luasnPhelylslystyrsnlystyrylleleuLys..... 610
1618 CCTCCGGAAGGTACGACCTAGAGTCATACACAAACTAGATGATCCAA 1667
611 .....AsnleuasplysphenleuasnmetSerSergluL 622
1668 ATGTCACAGGATAGAGATGATGATATATATATATACATATAA 1710
622 slysgluaspilleanserTyrllysasnlystyrglyleuLys 636

seq_name: pir2:T30933

seq_documentation_block:
chitinase (EC 3.2.1.14) A - Pseudocalteromonas sp. (strain S9)
C:Species: Pseudocalteromonas sp.
A:Variety: strain S9
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T30933
R:Techkanrananark, S.; Goodmap, A.E.
Microbiology 145, 925-934, 1999
A:title: Multiple genes involved in chitin degradation from the marine bacterium Pseu
A:Reference number: Z20935; M0ID:9923578
A:Accession: T30933
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-1054 <NEC>
A:Cross-references: EMBL:AF007894; NID:93928771; PID:93928775; PIDN:AAC79665.1
A:Gene: chIA
C:Keywords: glycosidase; hydrolase

alignment_scores:
Quality: 138.50 Length: 581
Ratio: 0.584 Gaps: 29
Percent Similarity: 40.792 Percent Identity: 19.105

alignment_block:
US-09-579-383-2 x T30933 ..

Align seg 1/1 to: T30933 from: 1 to: 1054

319 ATAGCAGTACTATAGTTGATGG...AACAGTCAGGTGATAGACAAA 365
322 ValileglytyrphenThrSerTrpragsnlyAlaasnlyglinProse 338
366 ACATATGATGATTCAAACCCAGATGGTGTCAATTTATATAT.....G 409
338 rTyrlleuValasprileProtyrAsplysilethrHisleleasnTyra 355
410 CATTTCGTCGCAAT..... 423
355 larpheAlaHisValasprAlaasnlyValSerileglyAspProSer 371
424 .....AATATGTTATAT..... 435
372 AlaAlaGlyasnProAlaThrAsnmetGluTrproGlyValAlaGlyAl 388
436 .....GATGTATCTAGACCATTTAATGAGAACAAAGATTCCTATTAA 478
388 agluMetaspProthrleuProtyrlysglyHis..... 399
479 GAAAACACGGCTTAGAATATGAACCTATGCTATGATGCTTAAATTAAT 528
400 .....PhasnleuL 403
529 AGAGTATCAGAAAGTACGTCCAGATGTAATTAATTTATTCCTTAGG 578
404 AsnlystyrylslyleuHisProaspVallystyThrleuileSerValGl 420

```

```

579 TGA..... 582
420 YGLYTRPAlaGluThrGlyGlyTyrPheAspAlaAsnGlyAsnArgValA 437
583 ..... GAACCTATATGATGATATGAAAAAGAAATTTGAT 618
437 laSerGlyGlyPheTyrThrMetThrThrAlaAspGlySerValAsn 453
619 ..... TATGGATAAATATGAGCTTGTAATGA 650
454 ThrAlaGlyIleAsnAlaPheAlaLysSerAlaValAlaGluPheIleGluTh 470
651 TTTTGATTTAGAGGTGATGATGATGAGGAA...CCACATGGGAAGT 697
470 rTyGlyPheAspGlyValAspIleAspTyrGluTyrPro..... 483
698 TTTACACTTAATGAATTAATTTTCAATTTATATTAATTAATTAAT 747
484 ..... SerSerMetAsnAspSerGlyHis..... 491
748 AACTGTTAAGAAAACTATCCGAGAAAGTAAATTTCAATTTCTG 797
492 ..... ProAspAsp..... PheProIleSerAs 499
798 TTCAATCAAAATGCGCATTTATCATCGTTTCAAGAGTTCATCTTCTGTA 847
499 nAlaArgArgAlaGlyLeuAsnAlaSerTyrGluValLeu..... 512
848 AAGATGAAGAATCTCCATATACACTAAATTTTGTGACAAATAGAA 897
513 ..... MetLysThrLeuArgGluGluLeuAsp 521
898 ACAAAATAAGAAATTTACATAGGAGCAGACGATGTA..... 933
522 LysAlaGluGluLeuAlaGlyLysHisIstYrMetLeuThrIleAlaSerPr 538
934 .TCAGCGAGACTTTTATTT..... AATATTTTAAATCAGCAAGG 973
538 oSerSerGlyTyrLeuLeuArgGlyMetGluThrPheGlnThrHis. 554
974 AGAAATAGATCTTTATTTCAACATATACATTTAGAAACAAACAAT 1023
555 ..TyrLeuAspTyrValAsnIleMetSerTyrAspLeuHisGlyAlaTrp 570
1024 CCAGAT..... 1029
571 AsnAspHisValGlyHisAsnAlaIleAlaLeuTyrAspThrGlyLeuAspSe 587
1030 ..... ATATGCGTATGATGATGATGATGATGATGATGATGATGAT 1072
587 rGluLeuAlaGluTrpAsnValTyrGlyThrLysGluPheGluGlyIleG 604
1073 AATATACATCACAATCATATGAGTTTTCATTTAGAACATACAGAGCT 1122
604 lTyTyrLeuAsnThr..... AspTrpAlaValAlaThrTyrPheArgGly 617
1123 GGATTAGT..... 1131
618 GlYLeuSerAlaGlyAlaGlyIleAsnIleGlyValProTyrTyrThrArgG 634
1131 ..... 1131
634 yPheLysAspValSerGlyGlyLysGluAsnGlyLeuTrpGlyArgAlaAla 651
1132 ..CCGCAAAATAAGAAATTTAGAAATTTGAGAAAAACAATCATGAT 1179
651 euProAsnGluAlaAspCysAlaLys..... GlyThrGlyValGlyGlu 665
1180 AAAAAACAATAATATATAGCGAGATGTATATA...GGGATATGCGATTT 1226
666 LysAsnLysCysGlyAsnGlyAlaValAlaGlyIleAspAsnMetTrpHisAs 682

```

```

1227 ATTATGAAA..... GAACAATTAACAGCTGATCATTCGATGTAGATA 1270
682 pleuAspGluAsnGlyAsnGluMetProAlaGlySerAsnProLeu...T 698
1271 TTTTCTTACAAATTTTGGAAACATTTAAATCCGAGAGTCAACATCCA 1320
698 rPhisValLysAsnLeuGlnAsnGlyIleLeuProSerTyrLeuAla 714
1321 AAAGCTTACTATTAAGTAAACCCGTGAGACTGTAGACACAAATAGATGA 1370
715 TyrglyLeuAspProAlaAsnAspProThrAsp..... 725
1371 ATATGTTCCAGAGCTGTTATTTCCACCATAGCGATATATACAAACACA 1420
726 ..... LysLeuValGlyThrTyrThrArgHisT 735
1421 ATGATGCTATA..... TGAACAACTAGATCTTATTTCA 1452
735 yrasSerValAlaValAlaProTrpLeuTrpAsnAlaAspLysAsnVa 751
1453 ATTCAATGACCTGGTGTAGACAGATATGAATGGAGCTGTGTCAAAGTAT 1501
751 lPheLeuSerIleGluAspGluGluSerMetGlyThrLysLeuAspTyrV 768
1501 ..... 1501
768 alIleAsnLysGlyLeuGlyIleMetPheTrpGluLeuAlaGlyAsp 784
1502 ..... GCTATGAAAAAATATGCGATGGAGAAAGCAGCCAT 1536
785 pheAspTyrAspSerAlaLysGlyGluTyrPheMetGlySer..... 798
1537 TATATTAACACTGACTATTAAGAAAGCTTATTTATATGGAAG... 1582
799 ..... ThrMetThr..... SerLeuAlaTyrAsnLysPheA 809
1583 ..... GCGAACCATATTTAATTAATGCTGGCAACAG 1615
809 laGlnAlaGlySerLysTyrAspIleAsnLysGlyGluLys 822

```

seq\_name: p1r2:B71612

```

seq_documentation_block:
hypothetical protein PF0555c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71612
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: B71612
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3973 <GAR>
A:Cross-references: GB:AE001402; GB:AE001362; NID:g3845213; PIDN:AC71900.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0555c
C:Superfamily: Plasmodium falciparum hypothetical protein PF0555c

```

```

alignment_scores:
Quality: 138.00 Length: 675
Ratio: 0.471 Gaps: 34
Percent Similarity: 43.407 Percent Identity: 20.889

```

alignment\_block:

us-09-579-383-2 x B71612 ..

Align seg 1/1 to: B71612 from: 1 to: 3973



```

3271 eAsnAsnAsnGlnTyrIleArgAsnLysAspHisIleIlePheCysSerL 3288
1542 TAACACTGACTATAAAGAAGCTCTATTATTATATGAAAGGGAACCAT 1591
      ::::: ||| ||| ||| ::::: |||
3288 euIleValPheIleIleLysLeuTyrValPheValLys..... 3301
1592 ATTATATTAATGTGGCAACAGACCTCCGGAAGGTCAGGCACATAGAG 1641
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3302 .....AsnPheLysHisLysAsnValAsnPheAsnAs 3312
1642 TCATACACAAACTAGATGCATCCAAATGTCACAGGATAGAAGATGAA 1691
      ||| ::::: ||| ||| ||| ||| |||
3312 nHisGlnAsnAsnLysAsnHisGlnAsnVal..... 3322
1692 TAAAAAATATCCACATAAACAC 1714
      ||| ::::: |||
3323 .....AsnThrAsnValGlnHis 3328

seq_name: p1r1:Q0UTC5

```

```

seq_documentation_block:
MADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion
C.Species: mitochondrion Trypanosoma brucei
C.Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-May-2000
C.Accession: A04519, D22845
R.Henriksen, U.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12, 7327-7344, 1984
A.Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift con-
ondrial maxi-circle DNA.
A.Reference number: A93537, M01D:85037915
A.Accession: A04519
A.Molecule type: DNA
A.Residues: 1-590 <HEND>
A.Cross-references: GB:M94286; NID:q343546
A.Note: this translation is not annotated in GenBank entry TRBKPGEN, release 109.0
C.Comment: The DNA sequence is from a segment of the 20-kb maxi-circle, which is believed
C.Genetics:
A.Genome: mitochondrion
A.Genetic code: SGC6
C.Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C.Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

```

```

alignment_scores:
      Quality: 136.50      Length: 661
      Ratio: 0.533      Gaps: 35
Percent Similarity: 38.729      Percent Identity: 19.062

```

```

alignment_block:
US-09-579-383-2/rev x Q0UTC5 ..

```

```

Align seg 1/1 to: Q0UTC5 from: 1 to: 590

1696 TTTTATTCATTCCTCTATCCCTGACATTTGGATCATCTAGTTTGTG 1647
||| ||| ::::: ||| ||| ||| ::::: ||| ||| ||| |||
25 PheLeuSerPheThrPheLeuSerLeuValMetIleIlePheIleValLeuCy 41
1646 TATG.....ACCTAGTGCCCTG..... 1630
||| ||| ::::: ||| |||
41 SMetIlePheSerPheLeuMetValSerVal.CysLeuTyrGlyTyrTyr 57
1629 .....ACCTCCGGAGGTCCTGTGTGCCACCATTTAATTAATGAGTT 1586
||| ||| ::::: ||| ||| ||| ||| ||| |||
58 TyrTyrAspPheCysLeuIleLeuMetLeuAspPheCysPheIleTyrPle 74
1585 CCCCTTCCATATAATATAGAGCTTCTTATGTCAGTGTATA...A 1539
: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
74 uThrTyrValCysSerGlyPheTyrMetPheIleMetLeuLeuIleAsnM 91
1538 TAAATGGGCTGCTTCCATGCATAT...TTTTCATAGCATACTTGAC 1492
::::: ||| ||| ::::: ||| ||| ||| ||| ||| |||
91 etValPheCysPheIleValPheTyrValPheTyrTyrMetTyrPheAsp 107

```

```

1491 CAA.....GTCCATTCATATCTGTCTACACAGGTGCAT 1457
::: ::::: ||| |||
108 MetLeuLeuGlyArgPheLeuIleIlePheThrIlePheValValCysMe 124
1456 GAATTGATTAACATCTAGTTTCCATAT..... 1429
||| ||| ::::: ||| ::::: |||
124 ValLeuPheIleLeuSerTyrAspPheLeuThrAlaTyrCysGlyTyrPg 141
1428 .....AGCATTCATGTGTTGTAATATATCCCTATGTTGGAATTAAC 1387
::::: ||| ||| ||| ||| |||
141 LeuLeuLeuGlyLeuPheSerPhePheLeuIleSerTyrPheTyrP..... 155
1386 GAGTCCTGGACATATATCATCTAT..... 1363
::::: |||
156 .....TyrArgPhePheAlaLeuLysPheGlyPheLysAlaPhePheI 170
1362 .....TGTCGTACAGTCTTCAGGTTTTCAGTTATAGTAAGTCTTTT 1320
::: ||| ::::: ||| ||| ||| ||| ||| ||| |||
170 eGlyLysIleGlyAspValLeuLeuIlePheAlaPheSerIleIlePheL 187
1319 GGAGTTTGTACTCTCAGATTTAATGTTTCCAAATAT.....TGT 1279
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 euSerAsnGlyPheCysMetThrThrPheTyrPheLeuAsnPhePheCys 203
1278 AAGAAAAATATCTACATGGAATGATCCAGTTGTAATGTTTC..... 1237
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 MetAspTyrTyrTyrIleGlyPheSerIleCysLeuLeuValGlyCysAl 220
1236 .TTTCATTAATTAATG..... 1222
||| ||| |||
220 aPheThrLysSerThrGlnPheGlyLeuHisIleTyrLeuProAspAlam 237
1221 .....SCATATCCCTATAC..... 1207
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 etGlnGlyProIleProValSerAlaLeuIleHisAlaIleThrLeuVal 253
1206 ATCTGCCCTATATATTTGATTTATATCATGATGTTTTCCTACCA 1157
::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 ValCysGlyIleIleLeuLeuSerPheValTyrTyrPysPheAspPheTr 270
1156 ATTCATTAATTAATTCCTT...ATTTCGGGACATAATTC..... 1123
||| ||| ::::: ||| ||| ||| ||| ||| ||| ||| |||
270 PheSerTyrPheTyrAsnLeuIleGlyTyrPheThrLeuIleLeuIleL 287
1122 .....ACCTGTGTATGTTCTAATGAAGAAACCTAATGATGATGAT 1078
||| ||| ::::: ||| ||| ||| ||| ||| ||| ||| |||
287 euMetThrLeuCysValPhe.....TyrAsnPheAspVal 298
1077 ATATTTTAACCAAAATATAATGAGATAGTACATATTCATCATTAAT 1028
::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 .....LysArgTyrVal 302
1027 CTGATTTGATGTTCTAATATGATGTTGAATTAATACAGATCATAT 978
: ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 L.....AlaPheSerThrIleCysGlnIleSerPheSerIlePheC 316
977 TTCTCTTGTGCTGATTAATAAATTAATAAAGTTCTGCTGATTAACAT 928
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 yscysLeuCysIleAspIleTyrIleGlySer..... 326
927 CGCTGTGCCCTATGATATCTTATTTGTTCTATTTGTTTCAGACAAA 878
::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 .....LeuPhePheCysTyrHisMetPheTyrLysAl 337
877 A.....TTAGTGT.....ATAGG..... 862
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 aThrLeuPheIleValIleGlyIleTyrIleHisIlePhePheGlyLeuG 354
861 .....AGATCTCATCTTTACAGAAAGATGACACTGCTCA 826
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 InsAspLeuArgCysTyrPhePheMetTyrPheCysGlyCys..ValLeuA 370
825 AACGCATGATTAATGACAGATTTGATGAACCAAGAAATTAATTAACCTTT 776

```

[illegible]







US-09-579-383-2 x CHIT\_BRUMA ..

Align seg 1/1 to: CHIT\_BRUMA from: 1 to: 504

```

313 GGTATTATAGCAGGA...TACTATGGTTCATGG.....AACG 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 G1YTYVALAAGGLYCSTYTYTRHASNTRPALAGLNTYATGASPL 38
348 TCAAGGTGATAGAGCAAAACATATGATGATCAAAACCATGGTCA 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 YGLUGLYLSPHELEUPROGLYASNLLEPROASNGLYLEUCSTYHISI 55
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 TTTATATATGTCATTTGCTGCATTAATATGTTATATGATGATCTGA 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 ILEUTYR...Alaphealalysvalasrgluleuglyasp...serlys 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448 CCATTATATGAGACAAAGATTCCTATTAGAAACACGGCTTGAATA 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 PROPHGLUTRPASNPSPLU.....AspTh 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
498 TGAACCTATGATGATGATGATTAATGAATTAAGACGTATCAGAAAGTAC 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 RGLUTRserlysglymetlyrseralavalthrlyseuargluthra 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 GTCCAGATGTAATTTATTTCTTTATCTTATAGTGGAACACCTAT..... 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 snprglyleulysvalleuSerlyrlyglylytyrasnphedlyser 111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 .....ATGATAGATATGAAAAAGAAATTTGATTTGATTAATATAT 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 Alallehethrglyllealalyseralagllysthrlygluthrphel 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 GAACCTTGTAATGATTT.....GATTAGATGGTGTAGATA 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 elysseralallealaphelarglyasnsnphaspjlypheaspl 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 TTGACTGGGA...CCACATGGGAAGTTTACAACTTAATGATTAAT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 eusprglylutyprvalgly..... 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 TTTCAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 Valalaglunhlsalalyseuvalglulamelysthralaphewa 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
771 GGAAGAAAGTAAATTTCAATTTCTGTTCACTCA.....AATG 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 lgluglu.....Alalysthrserglylysglnargleuleuthra 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
809 CTGCATATCATGCGTTTCAGAGTGCATCTTTCTGTAAGATGAAGA 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 lalalalaserlalglylysglythrllleasrglyserlyasnlvalglu 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
859 TCT.....CCATATACATAAATTTTGT..... 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 SerleuglylysasnphaspheuleuSerlyrlyasplleuhi 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
883 .....TCTGAACAATGAAGAAACAATTAAGATTAATGAGGAG 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 sglysertrglulysasnlvalasplleuhslylyseuhsprorhrl 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
923 CACCGATGTTATGACGAGAACTTTTATTAATTTTATTAATCA..... 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 ysglygluvalsergly.....lleglyllephasnthgluphe 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
967 .....GCAAGGAGAAATAGAGCT 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 Alalalasptrytrpalaserlysglymetrprolysglulyslleleil 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
987 T.....GTAATTTCAACATCAATTTTGAACCT..... 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 eglylleprmetlyrvalaglnglytrphrleuaspasnproserglut 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 .....ACAAAT 1023

```

```

281 hrAlalleglylAlalaserargproseralaserlysthrasn 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 CCAGATATTAATGTCATATATGAC.....TTATCCCATTAATATT 1064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 ProAlaglyglythrAlaserlytrpogluleCyslystlyleuysgl 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1065 TGGTTTAAATATTAATCAATCAATCATATATAGTTTTCATTAAGACATA 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 uglyglylysgluthrval.....HISG 322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1115 ACAGAGCT.....GCAATTAGTCCC 1134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 lngluclyvalaglylatyrmetyalysglyaspnltrptryglytyr 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1135 GAAATTAAGAACTTATTAATTTGATAGAAACAAATCATGATTAATAA 1184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 Aspasnngluluthrilletrglyleusmetlystrpleuylsgl 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1185 TCAAAATATATATATAGGAGATGATAGGATATAGCATTTA..... 1227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ytyrgly.....GlyAlapheliletrpalaleuasphrea 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1228 .....TTTATGAAAGACAAATTTACCACTGATCATTTGATGATGAT 1272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 spaspheThrcllysserCysglylysglyprotyrProleu..... 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1273 TTTCTCAAAATATTTGAAACATTTA.....AATCGTA 1307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 ...LeuasnlalleSerSerlguleugluclyglusergluasnprogl 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1308 ACTCAAACTCCAAA...GACCTTACTATTAAGTCA..... 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 ullethrthrngluclyproserillefthrcluthrcluthrlygluthra 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1342 ..AACCTGAAAGCTGTAGCACAAATGATGAATAT 1374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 spgluthrngluclythrsergluthrAlaIatyr 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: SwissProt_39:CHIA_SERMA
seq_documentation_block:
ID CHIA_SERMA STANDARD; PRT; 563 AA.
AC P07234; Q54275;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CHITINASE A PRECURSOR (EC 3.2.1.14).
GN CHIA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RA Koo J.C., Lim C.O., Choi Y.J., Kim C.Y., Baik J.D., Lee S.Y.,
RA Cho M.J.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 990 / QMB1466;
RA Jones J.D.G., Grady K.L., Suslow T.V., Bedbrook J.R.;
RT Isolation and characterization of genes encoding two chitinase
RT enzymes from Serratia marcescens.
RL EMBO J. 5:467-473(1986).
RN [3]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-31.
RC STRAIN=BJL200;
RA MEDLINE=95154677; PubMed=7851747;
RA Brumberg M.B., Eijlsink V.G.H., Nes I.F.;
RT Characterization of a chitinase gene (chia) from Serratia marcescens
RT BJL200 and one-step purification of the gene product.
RL FEMS Microbiol. Lett. 124:399-404(1994).

```

```

RN [4]
RP REVISIONS, AND X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=95219379; Pubmed=7704527;
RA Perrais A., Tews I., Dauter Z., Oppenheim A.B., Chet I., Wilson K.S.,
RA Vorgias C.E.;
RT "Crystal structure of a bacterial chitinase at 2.3-A resolution.";
RL Structure 2:1169-1180(1994).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01455; AAA26551.1; -
DR EMBL; X03657; CAA27292.1; -
DR EMBL; Z36294; CAA85291.1; -
DR PIR; A25090; A25090.
DR PDB; 1CTN; 20-JUL-95.
DR InterPro; IPR001579; Chitinase_2.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR000601; PKD_domain.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Hydrolyase; Glycosidase; Chitin degradation; Signal; 3D-structure.
FT SIGNAL 1 23
FT CHAIN 24 563 CHITINASE A.
FT DOMAIN 150 563 CATALYTIC.
FT ACT_SITE 315 315 PROTON DONOR (PROBABLE).
FT ACT_SITE 391 391 PROBABLE.
FT CONFLICT 52 52 N -> S (IN REF. 3).
FT CONFLICT 73 73 A -> T (IN REF. 3).
FT CONFLICT 76 77 TA -> GP (IN REF. 3).
FT CONFLICT 79 79 I -> V (IN REF. 3).
FT CONFLICT 121 121 T -> S (IN REF. 2).
FT CONFLICT 139 139 A -> P (IN REF. 2).
FT CONFLICT 226 226 V -> I (IN REF. 2).
FT CONFLICT 395 395 P -> A (IN REF. 1, 2 AND 3).
FT CONFLICT 410 430 PANKPTAYTVGVNALLAQ -> RPSRHRLHGERROC
FT CONFLICT 437 437 V -> I (IN REF. 1 AND 2).
FT CONFLICT 464 467 ATGP -> HRA (IN REF. 2).
FT CONFLICT 473 473 K -> E (IN REF. 2 AND 3).
FT CONFLICT 484 484 G -> S (IN REF. 2).
SQ SEQUENCE 563 AA; 60979 MW; 0696F6F6AF83AA35 CRC64;

alignment_scores:
Quality: 143.50 Length: 425
Ratio: 0.721 Gaps: 20
Percent Similarity: 46.824 Percent Identity: 20.471

alignment_block:
US-09-579-383-2 x CHIA_SERMA ..

Align seg 1/1 to: CHIA_SERMA from: 1 to: 563

25 TTAATTAATGATCATCATCTGATTCATTCGCAATTCAGAACCTTGAAAGG 74
||||:||||| ||| ||| ||||| ::| ::| ||
11 LeuLeuIleGlySerThrLeuGlySerAlaIleAlaIleAlaIleProG1 27
75 AAAAAAATAATATA.....AATAATTCATTTGGGAATAATACGGGAAA 115
||||| ::||| |||:|||||:|||||:|||||:|
27 yLysProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluValA 44
116 AATAAAATAAACAATCAACGGAATAATACGATGCTTTTTCACATCTT 165

```

```

44 spGlnAlaIleAlaThr.....AlaTrpAsnIleValLysVal 56
||||| ||| ::|||:|||||:|||||:|||||:|
166 AATACGAATAATAGTAATTTGTAGCAATATGAGATCTTTCGCGAGAT 213
||||:||||| ::| |||:|||||:|||||:|||||:|
57 LysAsnAlaIleAspValSerValSerTrpAsnLeuTrpAsnGlyAspAl 73
214 GGGTGAACCTCTAGAAATTACAAAAATAATATAA..... 246
||| ::|||:|||||:|||||:|||||:|
73 aGlyThrThrAlaLysIleLeuLeuAsnGlyLysGluAlaTrpSerGlyP 90
247 .....AATATAAATAAATATAT 264
|||||:|||||:|||||:|||||:|
90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyLys 106
265 AGA..... 267
|||
107 ArgTrpGlnMetGlnValAlaLeuCysAsnAlaAspGlyCysThrAlaSe 123
267 ..... 267
123 rAspAlaThrGluIleValAlaIleAspThrAspGlySerHisLeuAlaP 140
268 .....AATACCAACAAGACAATTTTAGAGGAGTATAAATAAAGCAACA 312
||||:||||| ::|||:|||||:|||||:|||||:|
140 roLeuLysGluPro.....LeuLeuGluLysAsnLysProTrpLysGln 154
313 GGT.....ATTATGACAGATACATGATGTCATGCAACAGTCAAGC 353
|||:|||||:|||||:|||||:|
155 AsnSerGlyLysValGlyValGlySerTrpPheValGlnTrpGlyValTyrG1 171
354 TGATAGACAAACAATATGATGATTCACAACCCAAATGGTGTCAATTTTAT 403
||| ::|||:|||||:|||||:|
171 Y.....ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThrH 186
404 ATATTCGATTTGCGCGCATT.....AATATGTAATGAT 438
||||:||||| ||| ::| |||
186 IseLeuLysGlyLysPheIleProIleCysGlyGlyAsnGlyLysAsnSP 202
439 .....GTATCTGACCATTTTAA 455
203 SerLeuLysGluIleGlyLysPheGlnAlaLeuGlnAlaGlySerCysG1 219
456 TGGAGACAAAGATCTCTATTAGAAAAC.....GGCTTAG 493
|||||:||||| ||| ::|
219 nGlyArgLysPheLysValSerIleHisAspProPheAlaIleLeuG 236
494 AATATGAACCATGATGATGATG..... 516
|||:|||||
236 InLysAlaGlnLysGlyValThrAlaTrpAspAspProTrpLysGlyAsn 252
517 CTATATGAATAATTAGCATGATGAGAAAGTACGTCAGATGTAATTATCT 566
||||:||||| ::|||:|||||:|||||:|||||:|
253 PheGlyGlnLeuMetAlaLeuLysGlnAlaHisProAspLeuLysIleLe 269
567 TTTATCTTAGTGAGGAAAC.....TATATGATAG 598
|||:||||| ||| ::|||
269 uProSerIleGlyLysTrpThrLeuSerAspProPhePheLeuMetGlyA 286
599 ATATGAAAAGAAATGATGATGATGATAAATATGACGCTTGATAT 648
|||:||||| ::|||:|||||:|||||:|||||:|
286 sPlysValLysArgAspArgPheValGlySerValLysGlnPheLeuGln 302
649 GATTTTGAT...TTAGATGCTGTAGATATTGACTGGGAA...CCAGATG 692
|||:|||||:|||||:|||||:|||||:|
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGlnPheProGlyG1 319
693 GAAGTTTAC.....AATTAATGAATTAATTTTCCAATATATATA 736
|||:|||||:|||||:|||||:|
319 yLysGlyLysAlaAsnProAsnLeuGlySerProGlnAspGlyLysThrTyv 336
737 TTAATATTAATTAATCTGTTAGAAAACATAATTCGCGAAGAAAGTTAAT 786
|||:|||||:|||||:|||||:|

```

```

336 allLeuMetlysgIleuAlaMetLeuAspGln..... 348
787 TCAATTTCTGGTTCATCAATAGTCGATTCATCGCTTTCAGAGTTGC 836
349 .....LeuSerAlaGluThcGly..... 354
837 ATCTTTCTGTAAGATGAGATCTCCATATACACTAAATT...TTGT 883
355 .....ArgLysTyrGluLeuT 360
884 CTGCAAAATAGAAACAAATTAAGATTAACATAGGCGAGCGCATGTTA 933
360 hrSerAlaIleSerAlaIleLysAspLysIleAspLysValAla..... 374
934 TCACGACGAACTTTTATTAATATTTTAAATACAGCAAGAGAAATAGA 983
375 .....TyrAsnValAlaIleAsnSerMetAs 383
984 TCTGTATTTATTCAAACATACAAAT 1008
383 PhisIlePheLeuMetSerTyrAsp 391

```

seq\_name: SwissProt\_39:NU5M\_TRYBB

seq\_documentation\_block:

```

ID NU5M_TRYBB STANDARD: PRT; 590 AA.
AC P04540;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN ND5.
OS Trypanosoma brucei brucei.
OG Mitochondrion.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OC NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037915; PubMed=6093040;
RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,
RA van Boom J.H., Benne R.;
RT "The sequence of the gene for cytochrome c oxidase subunit I, a
RT seven unassigned reading frames in Trypanosoma brucei mitochondrial
RT maxi-circle DNA.";
RL Nucleic Acids Res. 12:7327-7344(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=164;
RX MEDLINE=87201680; PubMed=2437452;
RA Jasmer D.P., Feagin J.E., Payne M., Stuart K.;
RT "Variation of G-rich mitochondrial transcripts among stocks of
RT Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 22:259-272(1987).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X01094; CAB57807.1; -.
CC EMBL: M14820; AAB59225.1; -.
CC PIR: A04519; Q0UTC5.
CC InterPro: IPR001750; Oxidored_q1.
CC Pfam: PF00361; Oxidored_q1; 1.
CC Oxidoreductase; NAD: Ubiquinone; Mitochondrion.
CC KW SEQUENCE 590 AA; 71494 MW; CAE231AD9DF5282 CRC64;

```

alignment\_scores: Quality: 136.50 Length: 661  
Ratio: 0.533 Gaps: 35  
Percent Similarity: 38.729 Percent Identity: 19.062

alignment\_block:

us-09-579-383-2/rev x NU5M\_TRYBB

Align seg 1/1 to: NU5M\_TRYBB from: 1 to: 590

```

1696 TTTTATTCATTTCTTATCCCTGACATTTGATGATCATGTTTGTG 1647
1696 TTTTATTCATTTCTTATCCCTGACATTTGATGATCATGTTTGTG 1647
25 PheLeuSerPheTyrLeuSerLeuValMetIlePheIleValLeuCy 41
1646 TATG.....ACTTAGTCCTG..... 1630
41 smetIlePheSerPheLeuMetValSerVal.CysLeuTyrGlyTyrTyr 57
1629 .....ACCTTCGGAGGCTCTTGTGCGCACCATTTAATAATATGTT 1586
58 TyrTyrAspPheCysLeuIleLeuMetLeuAspPheCysPheIleTyr 74
1585 CCCCTTCCATATATATATAGAGCTTCTTATAGTAGTGTTA...A 1539
74 uThrTyrValCysSerGlyPheTyrMetPheIleMetLeuLeuIleAsn 91
1538 TAATGGCTGCTTCCCATCGCATAT...TTTTCATAGCATACTTGGAC 1492
91 etValPheCysPheIleValPheTyrAlaPheTyrTyrMetTyrPheAsp 107
1491 CAA.....GTCCATTCATATCTGTGTACACAGGTCAT 1457
108 MetLeuLeuGlyArgPheLeuIleIlePheTyrIlePheValValCys 124
1456 GAATTCATATAGATCTAGTTTCCATAT..... 1429
124 tAsnLeuPheIleLeuSerTyrAspPheLeuThrAlaTyrCysGlyTyr 141
1428 .....ACCATCATGTTGTGTATATATATCCATGTTGGATTAAC 1387
141 IuLeuLeuGlyLeuPheSerPhePheLeuIleSerTyrPheTyrP... 155
1386 GAGTCTCGAACAATATTCATCTAT..... 1363
156 .....TyrArgPhePheAlaLeuLysPheGlyPheLysAlaPhePhe 170
1362 .....TGTCCTACAGTCTTCAAGGTTTTCAGTATAGAGTCTTT 1320
170 eGlyLysIleGlyAspValLeuLeuIlePheAlaPheSerIleIlePhe 187
1319 GAGTGTGTACTTCAGATTTAATGTTTCCAAATAT.....TGT 1279
187 eusSerAsnGlyPheCysMetThrPheTyrPheLeuAsnPhePheCys 203
1278 AAGAAAAATATCATACGATGATCCAGTTGTAATTTGTC..... 1237
204 MetAspTyrTyrTyrIleGluPheSerIleCysLeuLeuValGlyCysAl 220
1236 TTTCAATAAATAATG..... 1222
220 aPheThrLysSerThrGlnPheGlyLeuHisIleTyrLeuProAspAla 237
1221 .....CCATATCCCTATAC..... 1207
237 etGluGlyProIleProValSerAlaLeuIleHisAlaIleThrLeuVal 253
1206 ATCTGCCATTAATATATTTTGAATTTTATCATGATGTTTTCACACA 1157
254 ValCysGlyIleIleLeuLeuSerPheValTyrTyrCysPheAspPheTr 270
1156 ATTCTAATAATTTCTT...ATTTCGGAGACTAATATC..... 1123
270 pPheSerTyrPheTyrAsnLeuIleGlyTyrSerThrIleLeuIleL 287

```

```

1122 .....ACCTGTTATGTTCTAATGAAAAACCTAATGATGTTGATGTT 1078
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 eumethleucysValphe.....Tyrasnphaspal 298
1077 ATATTTAAACAAATATTAATGGATAGACATATCTACCATTAAT 1028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 .....Lysargtyva 302
1027 CTGATTTGACTTCTAATGTTGTAATTAATACAGATCTAAT 978
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 L.....AlapheserThrllecysglInIleSerPheSerMetPhec 316
977 TTCTCCTTGTCTGATTAATAAATTAATAAGTCTCTGATGATACAT 928
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 yScylaucysIleasplleTyrlleglyser..... 326
927 CGCTGCTGCCATGATATCTTATTTGTTCTATTTGTTTCAGACAAA 878
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 .....LeuphephecystyrylHismelPheTyrlLysal 337
877 A.....TTTACTTT.....ATATGC..... 862
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 athrlleupheleValleuglyleTrrpleHlsllePheheglyleug 354
861 .....AGATTTCTCATCTTTACAGAAAATGCAACTCTCGA 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 InAspleuargystyrylPhephecMetTyrlPhecysglycys..Valleua 370
825 AAGCATGATATATGACGATTTGATGACACAGAAATGAAATTAATCTTT 776
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 laarLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 386
775 CTTCGGATAGTTTCTTAACAAGTTAATTAATTAATAAATTAATTT 726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 Leu..... 387
725 GAAAAATTAATTCATTTAAGTTGTAACCTCCATGTCGTCACAGTC 676
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 .....Cysgly..... 389
675 AATATCTACACCATCTAATCAAAATCATTTACAGCTTCATATTTTAT 626
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
390 .....PheTyrl 391
625 CCACATTAATCATTTCTTTCTATATCTATCATATAGTTTCTCCACCT 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 CyslyAspmetLeuLeuAlaleuLeuMetLeuLeuSerPheTyrlasn.. 407
575 AAGGATAAAGATAATTAATCTGAGCTACTTTCTGATACGTCATAT 526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 .....llelleglupheleupheleleSerllelePhelellepe..... 420
525 TTCATTAGCATCATACATAGTAGTTTCATATCTTAAGCCGTTTCTTA 476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 .....PheThmetlleTyrlasnTyrlPheleu 429
475 ATAGGAATCTTTGCTCTCCATTAATGCTGATACATCATATAACATA 426
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 LeuphePheleuphelePhe.....Valph 437
425 TTAAATGCCAGAAATGCAATATATAAATGACACCATGCGTTTGAATC 376
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 eulysCys.....PhecysleuVala 444
375 AATCATATGTTTGTCTATACCTTGACTGTTCCATGACACATGATATC 326
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 spCysleupheleupheleupheasplTyrlgIucysleuValTyrlcysleu 460
325 CTGCTATAATAC.....CTTGTCTCTTTTATTAATCTCTCTAAAT 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
461 lleSerleuTyrlMetCyslleleuSerllelePhePhelele..lleasph 476

```

```

281 TGCTTTGCTG.....ATTTCATATCATTTTAT 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 eValCysllePheValPheSerSerTyrlCysValPheTrrPserPhePhele 493
252 TATATTTTATATTTTGTAAATCTAGATTAACACCATCTCCGCAT 203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
493 euAsnphetyrlasnPhe..... 498
202 AAGATCATATTTCTCAAAATTAATTAATTCATTAATTCATTAAGATGAAAA 153
498 ..... 498
152 GACTCATGATTTCCGTTGATGACTTTTATTTATTTTCCGATATAT 103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 .....PheaspllealalePheValalePheleulle 510
102 TCCCATGATATTTATAT.....TATTTTTC... 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
510 euSerValglPheleuTyrlTyrlgIucysleuphePheTyrlPheasn 526
73 .....CTTCAAGGTTCTGAAATTTG 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 lleaspcyslleMetleupheTrrPargile 536

```

seq\_name: SwissProt\_39:KTXA\_KLUOLA

```

seq_documentation_block:
ID KTXA_KLUOLA STANDARD; PRT; 1146 AA.
AC P09805;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE KILLER TOXIN ALPHA/BETA SUBUNITS PRECURSOR. (RP2 PROTEIN) [CONTAINS:
DE KILLER TOXIN ALPHA SUBUNIT; KILLER TOXIN BETA SUBUNIT (ENDOCYTINASE)
DE (EC 3.2.1.14)].
OS Kluyveromyces lactis (Yeast).
OC Plasmid PGK1-1.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_Taxid=28985;
RN [1]
RP SPRAIN-CBS 2359/152;
RA Sor F., Fukuhara H.;
RT "Structure of a linear plasmid of the yeast Kluyveromyces lactis;
RT compact organization of the killer genome.";
RL Curr. Genet. 9:147-155(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84297209; PubMed=6473099;
RA Stark M.J.R., Mlleham A.J., Romanos M.A., Boyd A.;
RT "Nucleotide sequence and transcription analysis of a linear DNA
RT plasmid associated with the killer character of the yeast
RT Kluyveromyces lactis.";
RL Nucleic Acids Res. 12:6011-6030(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037931; PubMed=6387625;
RA Tokunaga M., Kawamura A., Hishinuma F.;
RT "Cloning and nucleotide sequences of the linear DNA killer plasmids
RT from yeast.";
RL Nucleic Acids Res. 12:7581-7597(1984).
RN [4]
RP IDENTIFICATION OF PROTEIN, AND SEQUENCE OF 30-44 AND 895-916.
RX MEDLINE=87004569; PubMed=3758030;
RA Stark M.J.R., Boyd A.;
RT "The killer toxin of Kluyveromyces lactis: characterization of the
RT toxin subunits and identification of the genes which encode them.";
RL EMBO J. 5:1995-2002(1986).
RN [5]
RP SIMILARITY TO CHITINASE OF ALPHA-SUBUNIT.
RX MEDLINE=90259069; PubMed=2342564;
RA Bradshaw H.D. Jr.;

```









KW Transferase: DNA-directed RNA polymerase; Transcription; Zinc;  
 FT Zinc-finger; Nuclear protein. C3H-TYPE (POTENTIAL).  
 ZN-FING 88 101  
 SO SEQUENCE 2339 AA; 272829 MW; EDD99363086FD48 CRC64;

## alignment\_scores:

Quality: 131.00 Length: 531  
 Ratio: 0.524 Gaps: 27  
 Percent Similarity: 47.081 Percent Identity: 19.962

## alignment\_block:

US-09-579-383-2 x RPCL\_PLAFA ..

Align seg 1/1 to: RPCL\_PLAFA from: 1 to: 2339

```

55 AATTCGAGACCTTGAAAGGAAAAAATAATATAATATTCATTGGAGAT 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1533 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn... 1548
105 AATACGGGAAAAATAAAATAAATCAATCAAAACGGAATACATGACTCT 154
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1549 ... AsnAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsp 1564
155 TTTCACATCTTAATCGATATATAGTAATTTGTAGAAATATGATCTTAT 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1564 aAsnAsnLeuGlyValThrAsnTyrAsnThrAsnIleTyrProAsnAsp 1580
205 TCCGCGAGTGGGTGTAACCTGAAATTCACAAAATAATATAA..... 246
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1581 Cys... AsnGlyIleTyrGluLysGluThrAsnAsnAsnGluLeuThr 1596
246 ..... 246
1596 rAsnSerAsnMetCysAspLysAsnAsnAspPheSerAspGluPhePhe 1613
247 .. AATTAATAATAAAATGATAGAAAAATCCCAAGACAAATTTAGAGGAG 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1613 snAsnIleAsnGluAsnAsnAspLeuLeu.TyrAspAsnLysTyrTyrG 1629
295 TATTAATAAAAGGAAACAAAGTATTATAGCAGATACTATGTTTCATGAA 344
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1629 nIlePheLysAsn.....ValIleGlyP 1637
345 CAGTCAAGGTGATAGAG.....CAAAACATATGATG 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1637 heValSerValPheGluTyrValGluSerTyrLysGlnHisTyrIleLeu 1653
377 ATTCAA.....ACCAATGGTGTCAATTTTATATATTCGATTTGCT 417
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1654 PheProTyrGluIleIleLysThrPheSerPheLeuGluTyrIleLeu 1670
418 CCCATTAATATGTATATGATATGATATGACCATTAATG.....GAGG 461
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1670 rGluIleIleProThrAsnIlePheLeuHisThrLysLeuSerLysLysG 1687
462 ACAAAAGATTCCTATTAGAAAAACAGCGCTTAGAATATGAAACCTATGCTA 511
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1687 LuLysProThrHisGlnLysAsnThrGly..LysMetLysIleTyr.... 1701
512 TCATGCTTAATGAATTTAGACGTATCAGAAAAAGTACGTCACATGTAAT 561
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1702 .....IleGluIleLysLysTyrPheLeuHis.....AlaIle 1714
562 ATTCTTTTATCTTAGTGGAGAAACCTATATGATAGATATGAAAGAAAGA 611
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1715 AsnIleTyrLysTyrPheSerPheLysLysSerIleGluLeuIleLysL 1731
612 AATGATTATATGTGAATAAATATGTAAG..... 639
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1731 sLysAspTyrPheAsnTyrIleIleLysAsnTyrAspIleSerHisArgT 1748
640 ..CTTGTTATATGATTTTGATTTAGATGCTGTAGATATATTCACGGAACCA 687
```

```

1748 yTlleIleHisAspTyr..... 1753
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
688 CATGCGAGACTTTTACACTTAATATGATTAATTTTCAATATATATAT 737
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1754 .....SerPheIleAsnLeuLysGlnLeu.....TyrLeuP 1764
738 TAAATTAATTAATCTGTATAGAAAAACTATCCGGAAGAAAGCTTAATTT 787
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1764 eIlePhePheAsnIleTyrLysTyrPhe.....LysTyrIleS 1777
788 CAATTTCTGTTTCATCAATATCGTCATTTATCATGCTTTCA..... 828
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1777 eThrProGlyAspAlaValGlySerIleSerAlaGlnSerIleGlyGlu 1793
829 .....GGAGTTGTCATC 839
1794 ProGlyThrGlnMetThrLeuLysThrPheHisPheAlaGlyAlaIle 1810
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
840 TTTC.....TGTAAAGATGAAG 856
1810 rMetAsnValThrLeuGlyValProArgIleLysGluIleIleAsnAlas 1827
857 AATTCGCATATAACACTTAATTTTGTCTGACAAATATAGAAACAAATATA 906
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1827 eTrAsnSerIleGlnThrProIleLeuAsnIleProLeuGluValAsnAsp 1843
907 GAATTATACATGAGGACAGCAGCATG..... 930
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1844 AsnTyrAsnPheAlaLeuMetLysSerLysLeuGluLysThrTrl 1860
931 .....TTATCAGCAGAACTT 946
1860 eArgAspIleCysMetTyrIleLysGluAspTyrThrSerArgGlyValP 1877
947 TTATTAATATTT...TTAATACAGCAAGAGAGAAATATACATCTTGATTT 993
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1877 heLeuSerValLysPheAsn.....GluGluLeuIleGlnLysLeuPhe 1891
994 .....ATCAACATATACATTTTGAAGAACTACAAATCCAGATTAATGCT 1037
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1892 LeuAsnIleAsnAlaTyrAsnIleLysAspIle.....IleLeuL 1905
1038 AGATATGTACTTATCCCATTTTATTTGTTTAAATATTAACATCAACA 1087
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1905 sGlnSerHisIleAsnLysIleLysIleAsn..LysIleHisIleAsnV 1921
1088 TCATATATGAGTTTTCATTTAGAACATACACAGAGGTGATTTAGTCCCGAA 1137
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1921 aIleLeuAsnLysTyrLysLeuHis.....IleSerLeuLys 1932
1138 AATTAAGAAATTTATTAATGATTTGTAGAAAAACAATATCATATATAAATCA 1187
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1933 AsnAspGluPheIle..... 1937
1188 AATTAATAATAGGCGAGATGTATAGGATATGCGATTTATTTATGAAAG 1237
1938 .....PhePheGlnMetG 1942
1238 AACATATACCACTGATCATTCGATGTAGATATTTTCTTACAAATATTT 1287
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1942 LuSerLeuLysLysGlyLeuLeuAspLeuLeuIleTyrGlyAsp..... 1956
1288 TCGAAACATTTAATCTGAGTACAAACTCCAAAAGACCTACTACTATAC 1337
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1957 ...LysAspIleLysArgCysIleIleLysLysGluAspIleGluValTh 1972
1338 TGAAAACCTGGAAGAC...TGTAGCACAATATGATGAATAT 1374
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1972 rAspAsnGluAspGluIleCysAspAspMetAspGluTyr 1985
```

seq\_name: Swissprot\_39:PTPL\_DICDI



```

826 TCAGAGAGTTCATCTTCTGTAA.....GATGA 854
      ::::: ::::: :::::
257 GtlyglutThrArgAspLleThrcInlYrGlnlYrGluGlyTrpProAspH1 273
      ::::: ::::: :::::
855 AGAATCTCCATATAACACTAAATTTTGTCTGACAAATAGA.....A 898
      ::::: ::::: :::::
273 sasnIleProAspH1stHcInlProPheArgInlLeuHisSerIleT 290
      ::::: ::::: :::::
899 CAATAAAGAA..... 909
      ::::: :::::
290 hnsArGlnsncInlIleleProSerSerAspArgAsnValProIle 306
      ::::: :::::
910 ..TTACATAGGCAGACGCGATGTTATCAGCAGCAACTTTTATTAATAT 956
      ::::: :::::
307 lIeValHiscSerAlaGlyValGlyThrInlYrThncPheCysThrAl 323
      ::::: :::::
957 TTTTAAATACGAAGAAGAAATAGATCTTGTATTATTCACACATYCA 1006
      ::::: :::::
323 aValIleMetMetLys...LysIleuAspH1stYrPheLysGln..... 336
      ::::: :::::
1007 ATTGAAGAACTACAAATCCAGATATAAGTATGATATGATATCCCAT 1056
      ::::: :::::
337 ..LeuAspAlaThrProIleAspGlnValAlaAspProPhe...ThrHis 351
      ::::: :::::
1057 TTAATATTT.....GTTTAA 1073
      ::::: :::::
352 LeuProIleThrGluYrGlnSerAspAsnLeuAspLeuLysGlyLeuG1 368
      ::::: :::::
1074 ATATACATCACATCATATTAGTTTTCATTAGACATACAGAGGTG 1123
      ::::: :::::
368 YTrHsPheLys.....SerIleYrAsnSerAsnG 380
      ::::: :::::
1124 GATTAGTCCGAAATTAAGATTTATTAGATTCGTAGGAAACAAATA 1173
      ::::: :::::
380 lYtIleAsnAsnAsn.....AsnAsnAsnLeu 389
      ::::: :::::
1174 CATGATAAATAATCAAAATATAATAGGCAGATGCT 1209
      ::::: :::::
390 AsnAsnAsnAsnIleAsnAsnSerAsnGly 401
      ::::: :::::
seq_name: SwissProt_39:CHIT_NPVAC
seq_documentation_block:
ID CHIT_NPVAC STANDARD; PRT; 551 AA.
AC P41684;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94301173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
polyhedrosis virus";
RL Virology 202:586-605(1994).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC -1- N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L22858: AAA66756.1; -.
DR HSSP: P07254: 1CTN
DR InterPro: IPR001579: Chitinase_2.
DR InterPro: IPR000886: ER_target.
DR InterPro: IPR001223: Glyco_hydro_18.
DR InterPro: IPR000601: PKD_domain.
DR Pfam: PF00704: Glyco_hydro_18; 1.
DR SMART: SM00089: PKD; 1.
DR PROSITE: PS00014: ER_TARGET; 1.
DR PROSITE: PS01095: CHITINASE_18; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW Endoplasmic reticulum.
FT CHAIN 1..17 POTENTIAL.
FT ACT_SITE 305..305 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 173..173 N-LINKED (GLCNAC:..) (POTENTIAL).
FT CARBOHYD 444..444 N-LINKED (GLCNAC:..) (POTENTIAL).
FT SITE 548..551 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 551 AA; 61368 MW; 4DDAAD187873BBA2 CRC64;

alignment_scores:
Quality: 129.50 Length: 417
Ratio: 0.689 Gaps: 21
Percent Similarity: 45.084 Percent Identity: 19.424

alignment_block:
US-09-579-383-2 x CHIT_NPVAC ..
Align seg 1/1 to: CHIT_NPVAC from: 1 to: 551

286 TTAGAGGAGTATAA.....AAAGAAACAGGTATTATACGAGATA 329
      ::::: :::::
137 LeuGlnAsnAsnLysProGlyArgArgGlnAspLysIleValAlaAlaI 153
      ::::: :::::
330 CTATGTTTCATGAGAACACTCAAGTGTATGATAGACAAACATATGATGATT 379
      ::::: :::::
153 rPheValGluTrpGlyValYrGly.....ArgAsnHepProValAspL 168
      ::::: :::::
360 CAACCCAAATGGTGCATTTTATATATGCAATTTGCTGCATTT..... 423
      ::::: :::::
168 ySValProLeuProAsnLeuSerHisLeuLeuYrGlyPheIleProIle 184
      ::::: :::::
424 .....AATAGTT 431
      ::::: :::::
185 CysGlyGlyAspGlyIleAsnAspAlaLeuLysTrpIleProGlySerPh 201
      ::::: :::::
432 ATATGATGTATCTAGACCATTTTAATAGAACAAAGATTCCTATTAGAA 481
      ::::: :::::
201 eGluSerLeuGlnArgSerCysLysGlyArgGlnsPheLysValAlaI 218
      ::::: :::::
482 AACAC.....GGCTTAA... 495
      ::::: :::::
218 lEhIsAspProTrpAlaAlaValGlnLysProGlnLysGlyValSerAla 234
      ::::: :::::
496 ..TATGAACCTATGATGATGCTTATGAAATTAACGATGAGAA 542
      ::::: :::::
235 TrpAsnGluProTyrLysGlyAsnPheGlyGlnLeuMetAlaAlaLysLe 251
      ::::: :::::
543 AGTACGTCAGATGTAATATCTTTATTCCTTAGGTGAGAAC... 588
      ::::: :::::
251 uAlAsnProHisLeuLysIleLeuProSerIleGlyTrpThrLeus 268
      ::::: :::::
589 .....TATATGATAGTATAGAAAGAAATGATATATG 624
      ::::: :::::
268 eRAspProPheTyrPheMetHisAspValGluLysArgAsnValPheVal 284
      ::::: :::::
625 GATAAATATGAACCTGTTAATGATTTGAT...TAGATGGTGTGA 671
      ::::: :::::
285 AspSerValLysGluPheLeuGlnValTrpLysPhePheAspGlyValAs 301

```



```

236 AA.....AATATAAATAATATAAATAGATAAATCACCA 276
    |||:||||| |||:|||||
568 YSGLNHISILEASNGILYSALALIASNLYSASN.....SerPro 581
277 AGACAAATTTAGAGAGATATAAAGAA.....CAAG 314
    |||:||||| |||:|||||
582 GIUPHELLEULYSGLYASPRYSGLMETLUSLEUALAPHELEULYGI 598
315 TATATACAGAGATCTATGCTCATGAGACATCAAGGTATAGACCA 364
    |||:||||| |||:|||||
598 YLEULESERGLY.....ASPGLYTYRVALSERLYSAPRGLYARGVALG 613
365 AACATATGATTTGATTCAAACCCATGTGTCAATTTATATATTCATTT 414
    |||:||||| |||:|||||
613 INILETYRTHRSEGLUGLNULEULYSGLINLEUHSILEULEU 629
415 GCTGCATTAATATGTATATGATGATCTAGACCATTATATGGAAGCA 464
    |||:||||| |||:|||||
630 SERASPLEUGLYMETLETYSERILETHRYS..... 640
465 AAGATTCCTATTAGAAAACAGCGCTTAGAATATGAACSTAGTATGA 514
    |||:||||| |||:|||||
641 .....ILEYSGIUGIUGIUGIULYSILEGU.....ILEL 651
515 TCGTTAATGAATTAGACGTATCAGAAAAGTAGCTCAGATGTAATTAT 564
    |||:||||| |||:|||||
651 YSARGASNGIULIEVALARGASNTRYLSLEU..... 661
555 CTTTATTCCTTAGGTGGAAACCTATATGATATAGAAAA..... 609
    |||:||||| |||:|||||
662 .....TYRVALILEGLIULIEALYASNCY 670
609 ..... 609
670 STHGLUSPLEULYSPTYRVALILEPROLYSTYRYSLSGLUARGI 687
610 .....GAAATTCATTTATGCGATAAATATG... 636
687 IELYSPTIOALASNTYRASPINLEUPPTYRSPYRARGILELLEYS 703
637 .....AAGCTGTATGATTTGATTT 659
704 GLUHISLEUALGLYSILETHRASPLYSPTYRGLYASNASPTYR.... 718
660 AGATGCTGTAGATATTGACTGGGAACCATGAGAGTTTACAACTTAA 709
    |||:||||| |||:|||||
719 .....ALATRPYSSERASNANARGYLSLEULYSLEUA 730
710 ATGAATTA.....AATTTTCAAAATTATATATTAATTAATT 747
    |||:||||| |||:|||||
730 SNTHLLEUGIULYSILEGLUGLNULEUASNPROHISLEUALRGILUGI 746
748 AACTGTTAAGAAAACATATTCG.....GAAGAAAGTTAT 785
    |||:||||| |||:|||||
747 ASNLYSPHELYSLEUASNILEPROHEGILEULYSGLIULELYSGLI 763
786 TTCAATTTCTGCT.....TCATCAATGCTGCAT 814
    |||:||||| |||:|||||
763 EASPLYASNGIYTYRVALTYRASPLEUSERVALGLUSAPRNSGLUASN 780
815 TATCATGCGTTTCAGAGACTTCATCTTCTGTAAAGATGAAGAATCTCCA 864
    |||:||||| |||:|||||
780 HEILETHALATHRGILE.....LEUCYS..... 788
865 TATAACACTAAATTTTCTGACAAATATAGAAACAAT...AAGAAT 911
    |||:||||| |||:|||||
789 HISASNTHIRLEPHESETRILESMLEUGIULEPROGLUPH 805
912 ACATAGGCGACAGCATGTTATCAGACGAACTTTTATATATATTTT 961
    |||:||||| |||:|||||
805 ELEULYSASPLYSPROALIALVALILEALAGLYTHRTHARGLYTHR 822

```

```

962 ATACAGCAAGAGCAAAATACATCTTGATTATTCAAAACATACAAATTA 1011
    |||:||||| |||:|||||
822 IASPTYRGLUGIUGIULALYLSLEULEU..... 832
1012 GAAACTACAAATCCAGATATATGATATGATATGATATCCATTATTA 1061
833 .....GLUALALEUALASPYALIMETMETGLUGIYASPAI 844
1062 TTTGGTTTAAATATAACATACACATATATAGTTTTCATTAGAAC 1111
    |||:||||| |||:|||||
844 AMEGLYYSRPROHE.....LEUPHEPROASNHELLEULEYSL 858
1112 ATACAGAGGTGATTTAGTCCGAAAATAAGAATTATTA..... 1152
858 EUNARGIULASNALAPHELYSASPRULASNLYSGIULEMETTYRYSILE 874
1153 ..GAATGTAGGAAA.....ACAAATCA 1175
875 HISGLINLEUSERALALYSPHEGLYILEPROTYRPHELLEASNMETLEUP 891
1176 TGATATAAATCAAAATATATATAGGCGAGATGTATA..... 1212
    |||:||||| |||:|||||
891 OASPTRGLINVALTHRASNTHRASNALAMETGLYCSARGYTHRARGLEUS 908
1213 ..GGATATGGCATTTATTTATGAAGACAAATTAACACTGATCATTC 1260
    |||:||||| |||:|||||
908 ERGLYASNTRP.....THGLYASPALA 915
1261 GATGATATATTTTCTTACAAT.....ATTGGAACAATTAATACCC 1304
    |||:||||| |||:|||||
916 GIULIEASPTHLLEUARGTHRGILYASNMETGLINTRPYRSELEASN. 931
1305 TGAAGTACAACTCCAAAAGACCTTACTATACGAAACCCGTGAAGCT 1354
    |||:||||| |||:|||||
932 .....LEUPROARG...ILEALTYRGLUALASNGLYASPAPT 944
1355 GTAGCACATATGATGATATGTTCCAGAGCTGCTATTCACCAACATAGG 1404
    |||:||||| |||:|||||
944 HRLYSLEUPHEGLULEULEUHSIGLUARGIULEULEULYSGLUALA 960
1405 ATATATTTCAAAACAAATGATGCTTATGAAACCTAGATCTTAT 1449
    |||:||||| |||:|||||
961 LEULEULEYSHISGLU.....VALTHRYSGIUALARGLEUTYR 973

seq_name: SwissProt_39:RPOB_PLAFA
seq_documentation_block:
ID RPOB_PLAFA STANDARD; PRT; 1024 AA.
AC P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RT "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
RL Plasmodium falciparum.";
Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT "A circular DNA in malarial parasites encodes an RNA polymerase like
RL Mol. Biochem. Parasitol. 44:115-124(1991).

```

```

CC      -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC      OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC      SUBSTRATES.
CC      -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE +
CC      RNA(N).
CC      -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC      -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X75544; CAA53332.1; -.
DR      EMBL; X52177; CAA36427.1; -.
DR      EMBL; X95275; CAA64572.1; -.
DR      PIR; S10438; RNZOF.
DR      InterPro; IPR001572; RNA_POL_B.
DR      Pfam; PF00562; RNA_POL_B.1.
DR      ProSite; PS01166; RNA_POL_BETA.1.
KW      Transference: Transcription; DNA-directed RNA polymerase.
SO      SEQUENCE 1024 AA; 122185 MW; B49BBD5D7C9BD412 CRC64;

```

```
alignment_scores:
  quality: 121.50
  ratio: 0.406
  percent_similarity: 45.929
  length: 651
  gaps: 37
  percent_identity: 20.737
```

```
alignment_block:
```

US-09-579-383-2/rev x RPOB\_PLAFA . . .

Align seg 1/1 to: RPOB\_PLAFA from: 1 to: 1024

[illegible]

```

391 .....:|||||:|||||:|||||:
581 CCACCTAAGATAAAGATATATATACATGACCTGTTTCTGATACG 532
400 L.....:|||||:
531 TCATATTCATTAAGCATACATACATAG.....TTTCATATTCATAGC 488
408 .....:|||||:
487 CGCTGTTTCTATAGATCTTCTCTTCATTAATAGCTAGATACCA 438
418 AsnIlePhe..AsnLysAsnPhe.....:|||||:
437 TCATATTCATTAATGCGAGCAATATATTAATTAATGACCATC 388
425 .....TyrAsnIleSerPhe.....AsnAsnIleTyrLeuLysAsnIle 438
387 TGGCTTGAATCATCATATGCTTTGCTATACCTTGAC..... 347
438 eAsnPhe..AsnLysThrThrIleLeuThrIleAsnLysAsnThrPheLys 454
455 IleCysAsnIleThrGlnAsnIleIle.....:|||||:
299 TTATATCTCTCTAAATTTGCTGCTGATTTTCATCATTTTATTTAT 250
464 TyrIleProPheAsnTyrLeuLeuSerPheIleGlnAsnLeuIlePro 480
249 ATTTTATATATTTTGTATATTCATAGACTACACCATCTCCGCAATAG 200
480 heIleHisTyr.....:|||||:
199 ATCCATATTCACAAATTTACATATTCATTAAGATGGAAGAAC 150
491 MetSerIleLysMetHis.....:|||||:
149 TCATGATATTCGCTTGCATGATTTTATTTTATTTTCCGCTATTTATCC 100
497 rGlnIleValPro.....:|||||:
99 CAATGATATTTATATTTTTCCTTCAAGCTTCTGCAAT 56
509 snIleIleThrAsnTyrAsnPheIleLeuAsnLysTyrLeuAsn 523

seq_name: SwissProt_39.YXDM_BACSU
seq_documentation_block:
ID YXDM_BACSU STANDARD; PRT; 622 AA.
AC P42424;
DT 01-NOV-1995 (Rel. 32, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHELICAL 70.5 KDA PROTEIN IN IDH 3 REGION.
GN YXDM OR B656.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC11;
RX MEDLINE=97021444; PubMed=8867804;
RA Yoshida K.-I., Fujimura M., Yanai N., Fujita Y.;
RT Cloning and sequencing of a 23-kb region of the Bacillus subtilis
RL genome between the iol and hut operons.;
RL DNA Res. 2:295-301(1995).
RN [2]
RP SEQUENCE OF 1-404 FROM N.A.
RC STRAIN=168 / BGSC11;
RX MEDLINE=95039891; PubMed=7952181;

```

```

RA Yoshida K.-I., Sano H., Miwa Y., Ogasawara N., Fujita Y.;
RT Cloning and nucleotide sequencing of a 15 kb region of the Bacillus
RT subtilis genome containing the iol operon.;
RL Microbiology 140:2289-2298(1994).
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; DA5912; BAA08316.1; -
CC EMBL; D14399; BAA03303.1; -
CC EMBL; Z99124; CAB15999.1; -
CC Subtilisin; BG1129; yxdm.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 154 174 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 498 518 POTENTIAL.
FT TRANSMEM 558 578 POTENTIAL.
FT TRANSMEM 590 610 POTENTIAL.
SQ SEQUENCE 622 AA; 70528 MW; B2FAFAD30D27370E CRC64;

alignment_scores:
Quality: 121.00 Length: 397
Ratio: 0.602 Gaps: 15
Percent Similarity: 50.630 Percent Identity: 19.899

alignment_block:
US-09-579-383-2 x YXDM_BACSU ..
Align seg 1/1 to: YXDM_BACSU from: 1 to: 622

505 TATGATATGATGCTTATATGATTAATGACGATACGAAAGATACGTCACA 554
|||||:|||||:|||||:
211 TyrGlyMetValLeuLys.....GlyAsnValHisGlyAlaGluProph 225
555 TGTATATATCTTTATCTTATAGTGAGGAGAAACCTATGATATATATAG 604
:::|||||:|||||:|||||:
225 eIleIleLeuLeuLeuThrValIleGly...ThrTyrPhePheSerG 241
605 AAAAGAAATTTGATTTATGATTAATATATGACCTTGTATATGATTTT 654
:::|||||:|||||:|||||:
241 InSerSerIle...TrpIleLeuArgAlaLeuLysTyrTrpLysThrPhe 256
655 GATTAGATGCTGTATATGATGACGGAACACATCGGAAGCTTTTACAA 704
||| ||| ||| ||| |||
257 TyrLeuArgGlyLysAsnIleLeuTrp..... 265
705 CTTAATGATTAATTAATTTTCAAAATTTATATTAATTAATTAATTAAT 754
::: ||| ||| ||| |||
266 .....ValSerAspLeuV 270
755 TAAGAAAACATATTCGAGAGAAAGTAAATTCATTTCTGCTTCATCA 804
::: ||| ||| ||| |||
270 AlTyrArgLeuLysAspAsnAlaArgLeuPhePheIleValSerIleIle 286
805 AATGCTGATATATCATGCGTTTCAGAGACTTGCACTTTCTGTAAGAT.. 852
:::|||||:|||||:|||||:
287 SerAlaValAlaIlePheThrAlaThrGlyValLeuAlaMetTyrLysSerTh 303
853 .....CAAGATCTCCATATATACATTAATTTTGTCTGCAACAA 892
|||||:|||||:|||||:|||||:
303 rValGlyAlaGluGlnSerAlaTyrGlnMetClnTyrLeuSerTyrSera 320

```

```

893 TAGAACAATAAAGATTACATAGGCGAGCGATGTTATCAGCAGCA 942
320 snasProlsGlnGlnThrHisLeuLysAspLLeuAspLLeuLys 336
943 ACTTTTATTAATTTTATACAGCAAGGAAATGATCTTAT 992
337 Thr.....HisGlyPheThrThrLysAspLysLLeuAspLLeuLys 351
993 TATTCACACATACATTTAGAACTACAAATCCAGATATATAGTGATGA 1042
351 ValAlaArgLysGlnGlnGlnGlnValProProValLysMetIleSerG 368
1043 TGTACTTATCCCATTTATATTT.....GCTTAAATAT 1077
368 LuserAspAlaLalLysLysPheHisValLysValAsnLLeuLysGln 384
1078 AACATCACAATCATATTAGTTTTCATTTAGAACATTAACAGAGGCGAT 1127
385 AspGlnAlaValLysPheProGlyThrLysAspArgAsnLLeuLysAsn 401
1128 TAGTCCGCAATAAAGATTATAGAAATGCTAGCAAAACAAATACATG 1177
401 uAlaProAspGlnLLeuLysLLeuAsnGlnLysGlnLLeuSerAspG 418
1178 ATAAAAATCAAAATATATATAGCGCATGCTATAGCATATGCAATTA 1227
418 InLys.....Leu 420
1228 TTTTGAAGAA...CAATTACCACTGCATCATCGATGATGATTTT 1274
421 SerValLysGlnValGlnLysProLLeuLysSerLLeuAlaIleLle 437
1275 TCTTCAAAAT..... 1284
437 aValAsnAspGlnThrPheAspGlnLLeuLysSerLLeuLysAlas 454
1285 .....ATTGGAACATTTAAATCCGAGGTACCAACTCCA 1320
454 erLeuTYrGlyTYrSerTYrAspLysPheLysAspSerLLeuLLeSer 470
1321 AAAGACCTTACT.....ATACTGAAAAACCTGGA 1349
471 GlnSerLLeuGlnAsnGlnLLeuTYrGlnAsnTYrLLeuAspValHis 487
1350 AGACTGTAGCAATATAGTGAATGTTCCAGACTGTTATCCACCA 1399
487 pPheAlaSerLysAlaGlyThrTYrTYrAspThrValGlnLLeuPro 504
1400 TAGGATATAT.....TACAACACATGATGCTATATGAAACT 1440
504 euserLLeuPheLLeuGlyLLeuPheLLeuGlnCysLLeuLysCys 520
1441 AGATCTTATTCATTCATGACCTGCTGCTAGACATATGATGAGGACT 1490
521 ThrSerPheSer..TYrPheArgLLeuPhe...ThrAspLLeuSpGln 536
1491 GGTCAAGTAT..GCTATGAAAAAATATGCGATGGGAACAGCCACTT 1537
536 rglLualTYrArgSerLLeuAlaLysLLeuGlyLLeuSerGlnArgGln 552
1538 ATTATACACTGACTATAAGAAAGCTCTATTATATAT 1576
553 AlagLInSerValThrLLeuGlnLLeuAlaLLeuPhePhe 565
seq_name: SwissProt_39:CHIT_NPVOP
seq_documentation_block:
ID CHIT_NPVOP STANDARD; PRT; 550 AA.
AC O10563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

```

DE PROBABLE ENDOCHITINASE PRECURSOR (EC 3.2.1.14).
OS Orygia pseudotsugata multicausid polyhedrosis virus (Opomyv).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RM [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SUBCELLULAR LOCATION: ENDOPASMIC RETICULUM LUMEN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: U75930; AAC59123.1; -.
DR HSP: P07254; ICTN.
DR InterPro: IPR001579; Chitinase_2.
DR InterPro: IPR000866; ER_target.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR000601; PKD_domain.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR SMART: SM00089; PKD; 1.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW Endoplasmic reticulum.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 550 PROBABLE ENDOCHITINASE.
FT ACT_SITE 304 304 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 146 146 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC...) (POTENTIAL).
FT SITE 547 550 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 550 AA; 60733 MW; 77947F5CF0E07BD CRC64;

alignment_scores:
Quality: 120.00 Length: 292
Ratio: 0.882 Gaps: 13
Percent Similarity: 46.575 Percent Identity: 21.233

alignment_block:
US-09-579-383-2 x CHIT_NPVOP ..
Align seg 1/1 to: CHIT_NPVOP from: 1 to: 550
319 ATACAGATCTATGTTGATGACAGCAAGTATGATGACCAACA 368
149 ValAlaLysPheValGlnThrGlyValTYrGly.....ArgLysPhe 163
369 TATGATGATTCACAAACCAATGCTGCAATTTATATATGATTCGTC 418
163 eProValAspLysValProLLeuProAsnLLeuSerHisLLeuLysCly 180
419 GCATTATATG.....TTATATGAT..... 438
180 helLeProLLeuCysGlyLysPheLysAspLysPheLysPheLysPhe 196
439 .....GATCTAGACCATTTAAAGAAAGCAAGCAAT 470

```



```

197 ProGlySerPheGluAlaLeuGlnArgSerCysLysGlyArgAlaAsp 213
471 CCTATTAAAGAAAACAC.....G 487
213 elysValAlaIleHisAspProTrrPalaIleGlnLysProGlnLysG 230
468 GGTTCGAA.....TATGAACCTATGGTATGATGCTTAATGAATTTGA 531
230 llyValSerAlaTrrPasnGluProTrrLysGlyAsnPhelGlyGlnLeu 246
532 CGTAAAGAAAAGTACGTCAGATGATTAATTTATCTTATCTTATAGTGG 581
247 AlaAlaLysLeuAlaAsnProHisLeuLysIleLeuProSerIleGlyG 263
562 AGAAAC.....TATATGATGATATAGAAAAAGAA 613
263 lYrrPheLysSerAspProPheTrrPheMetHisAspAlaLysArg 280
614 TTGATTATGTGATTAATTTGAAGCTTGTTAATGATTTTGCAT...TTA 660
280 rGvalPheValGlnSerValLysGlnPheLeuGlnValTrrPhePhe 296
661 GATGGTGTGATATGACTGGGA...CCACATGGAGATTTTACACTT 707
297 AspGlyValAspIleAspTrrPgluPheProGlyLysGlyAlaAsnPr 313
708 A.....AATGAATTAATTTTCAATTAATTTATTAATTAATTA 748
313 AlaLeuGlyAsnGlyLysGlnArgAspAlaAspTrrLysLeuValLeuL 330
749 ACTTTTAAAGAAACTATTCGCGAAGAAAGTTAATTTCAATTTCTGCT 798
330 ySglnLeuArgAlaMetLeuAspGlu..... 338
799 TCATCAATGCTGCATTAATCATGCGTTTCAGAGATTGCATCTTCTGTA 848
338 ..... 338
849 AGATGAAGATCTCCATATAACACTMAATTTTGTCTGAACAATAGAA 898
339 .....LeuGlnLeuGlnT 343
899 CAATTAAGATTAATGATAGCGCAGCGATGTTATCGACGAGAACTTT 948
343 hrAsnLysTrrTyrGlnLeuThrSerAlaIleSerSerGlyTrrAspLys 359
949 ATTATATTT.....TTATATACAGCAAGCAAAATAGATCTGTAT 992
360 lIleAlaValValLysTrrAspAlaIleGlnArgPheLeuAspLysIlePh 376
993 TATTCAAACATACAAATTTAGAA.....ACTACAATCCAGATATATATG 1036
376 eleuMetSerTrrAspPheLysGlyAlaTrrSerAsnThrAspLeu.... 391
1037 TAGATATGTACTATCCATTTATAT 1062
392 .....GlyTrrGlnThrThrLeuTrr 398
seq_name: SwissProt_39:HEMK_RICPR
seq_documentation_block:
ID      HEMK_RICPR      STANDARD;      PRT;      518 AA.
AC      G9ZCB3;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HEMK PROTEIN HOMOLOG.
GN      HEK OR RP847..
OS      Rickettsia prowazekii.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxID=782;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN-MADRID E.
RX      MEDLINE=99039499; PubMed=9823893;
RA      Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA      Sichteritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA      Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT      "The genome sequence of Rickettsia prowazekii and the origin of
RT      mitochondria".
RL      Nature 396:133-140(1998).
CC      -1- FUNCTION: PROBABLE METHYLTRANSFERASE.
CC      -1- SIMILARITY: IN THE N-TERMINAL REGION; BELONGS TO THE HEK FAMILY
CC      OF MODIFICATION METHYLASES.
CC      -1- SIMILARITY: IN THE C-TERMINAL REGION; BELONGS TO THE UPF0155
CC      FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AJ235273; CAI15271.1; -
DR      InterPro: IPR003358; Methyltransf_4.
DR      InterPro: IPR002052; N6_Mtase.
DR      InterPro: IPR000051; SAM_bind.
DR      InterPro: IPR000818; TEA.
DR      Pfam: PF02390; Methyltransf_4; 1.
DR      SMART: SM00426; TEA; 1.
DR      PROSITE: PS00092; N6_MTASE; UNKNOWN_1.
KW      Transferase; Methyltransferase; Complete proteome.
FT      DOMAIN 1 300
FT      DOMAIN 301 518
FT      FT 518
SQ      SEQUENCE 518 AA; 59564 MW; C66BB8DB273704A1 CRC64;

```

```

alignment_scores:
Quality: 119.00      Length: 422
Ratio: 0.583      Gaps: 25
Percent Similarity: 48.341      Percent Identity: 21.564

```

alignment\_block:

US-09-579-383-2 x HEK\_RICPR ..

```

Align seg 1/1 to: HEK_RICPR from: 1 to: 518
97 TTGGCAATTAATTCGGGAAATTAATTAATTAATCTCATCAACGGAATACA 146
   ::::::::::: ::::::::::: :::::::::::
107 lIleGlyLeuVal...ValSerArgAsnAsnLeuHisMet..... 118
147 TGAGCTTTTTCACATCTTAATCAATATAGT..... 180
   ||||| ||||| ||||| |||||
119 .....PheSerLysLeuLysSerLeuAspSerValLeuThrThrGlns 133
181 .....AATTTGTAGATATGAGATCTTATTCGCGAGATGGGTGT..... 219
   ||::::::::||| ||::: ||::: ||::: ||::: ||::: ||:::
133 ertyAsnIleLeuGlnLeuGlyThr.....GlySerGlyCysIleAla 147
220 .....AACTTAGAATTACAAAAAATTAATA 245
148 lIleSerLeuLeuCysGlnLeuProAsnThrAsnIleAlaIleAlaThrAsp 164
246 AAATTAATAATAAATGATAGAAATCAACAGACAAATTTTAGAGAGT 295
   ::::::::::: ::::::::::: :::::::::::
164  eSerValAspAlaIleLysValAlaLysSerAsnSerIleLysTrrAsnV 181
296 ATAAAAAAGCAACACAGCTATTATAGCAGATCACTACTATGTTCAATGAC 345
   ||| ||| ||| |||
181 alThrAspArgIleGlnIleIle.....HisSerAsnThrPhe 193
346 AGTCAAGCTGATAGCAAAA...CATATGATTGATTAACCCCA..... 387
   :::::::::: |||

```





ID RPOP\_AGABT STANDARD; PRT; 1102 AA.  
AC P33539;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).  
OS Agaricus bisporus.  
OS Mitochondrion.  
OG Plasmid pEM.  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Agaricales; Agaricaceae; Agaricus.  
OC NCBI\_TaxID=5343;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 24666 / AG4;  
RX MEDLINE=91347410; PubMed=1879001;  
RA Robison M.M., Royer J.C., Horgen P.A.;  
RT "Homology between mitochondrial DNA of Agaricus bisporus and an  
RT internal portion of a linear mitochondrial plasmid of Agaricus  
RT bisporus.";  
RL Curr. Genet. 19:495-502(1991).  
CC -! FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -! CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
CC RNA(N).  
CC -! SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
CC -----  
DR EMBL; X63075; CAA44799.1; -.  
DR PIR; S28104; S28104.  
DR InterPro: IPR002092; RNA\_pol\_phage.  
DR Pfam: PF00940; RNA\_pol\_1.  
DR PROSITE: PS00489; RNA\_POL\_PHAGE\_2; 1.  
DR PROSITE: PS00900; RNA\_POL\_PHAGE\_1; 1.  
KW transferase; transcription; DNA-directed RNA polymerase;  
KW Mitochondrion; Plasmid.  
FT ACT SITE 734 734 BY SIMILARITY.  
FT ACT SITE 804 804 BY SIMILARITY.  
FT ACT SITE 980 980 BY SIMILARITY.  
SO SEQUENCE 1102 AA; 126664 MW; F72FDD308D1AB5B4 CRC64;

alignment\_scores:  
Quality: 116.50 Length: 463  
Ratio: 0.537 Gaps: 21  
Percent Similarity: 46.868 Percent Identity: 21.166

alignment\_block:

US-09-579-383-2 x RPOP\_AGABT ..

Align seg 1/1 to: RPOP\_AGABT from: 1 to: 1102

```

10 AAAATGATCATATTTTAAATTAAGTATTCATCTTGATCTGCAAAATTC 59
   ||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
184 LysGluSerValValAlaIalysValGluGluIleasnTyrgluValHisAs 200
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
60 CAGAACCTTGAAGGAAAGAAATAATATAAT..... 90
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
200 nlyspheilelysglyasnTyrgluValHisAsnpheserAsnArgAsnLeut 217
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
91 .....AATTCATGGGAATATA... 108
   |||::: |||::: |||::: |||::: |||::: |||::: |||:::
217 hrcIuilelyserIleleuIyGluIleuAsnlysmetGluIleleuAsp 233

```

```

109 ...CGGGAATAATAAATAAATTAACATCAACGGAATAATAGTCTTT 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 AsnArgIleAsnlyserThrlyGluSerAspLeuIyValI 250
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
156 TTCACATCTTAATCGAATAT..... 177
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
250 elyGluIleleuAspserAsnleuIleIleGluAspLyGluIleuAlaI 267
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
178 .....AGTAATTTGTAGAAATATGA...TCTATTCGGAGATGGGTGT 219
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 IeGluIyThrValValGluIyThrIleuThrPhePheArgHisAsnMet 283
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
220 AACTCTAGATTAACAAAAATAT.....AAAATTAATAATAAATGA 263
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 AspThrIleGluThrArgAsnlyIleIleHisAsnIleTyrgly 300
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
264 TAGAAATACCAACAGCAAAATTTAGAGAGATTA... 300
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 uasnlyAlaTyThrIleuIleuAlaAsnTyrglyLeuAsnArgTyS 317
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
301 ...AAAGCAACAGATTAATAGCAGATTAATAGGTCATGCAACAGT 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
317 erlyIlelyserIle.....HisleuIleSerAsnlySer 330
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
349 CAACCTGATAGCAAAACATATGATTAATCAACCAATGGTGCAT 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 GluIyThrlyserIyGluIleIlelyleuIleValIleuValI 347
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
399 TTTATATAT.....GCATTTGTCGCATTA 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 elerlyIlelyIleAspIyScylleSerTyrglyThrIle 364
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
425 ATATGATATATGATGATATAGACATTTATAGACACAAAGATCTTA 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 IeAsnleuIleThrAsnAlaArgspIyThrSerArgThrAsnIleAla 380
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
475 TTAAGAAACAGCGCTTGAATATGAACCTATGATATG... 516
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 Ile.....AsnleuGlyPheArgIleIlelyValIleuIyTyrl 394
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
517 .....CTTATGAATTTAGACGATGCAAG 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 elyLeuAspIyIuAsnProSerIleAsnAlaIleTyrglyIleAsnlyS 411
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
545 TACCTCCAGATGATATTTCTTTATCTTAGCT.....GGA 582
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
411 eulysAspGluIleSerIyLeuAspAsnGluIyIyTyrglyIleGly 427
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
583 GAACCTATATATGATATGAAAAAGAAATGATATATGATATAAT 632
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 AspThrIleuGlyIleuIleThrAlaAsnIyAspIleValIleGluI 444
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
633 ATGAGAGCTGTTATGATTTGATTTAGATGAGTATGATGACGGG 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 uleuIyThrAsnSer.....GlyIyAspSerGluIleuG 456
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
683 AACCAATGGAAGTTTACACCTTAATGAATTAATTTTCAAAATTA 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 lu.....ValArgIleAsnAspIyPheIleSerAsnleu 467
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
733 TATATTAATTAATTAATGTTTACGA...AAACATATCCGAGACAAA 779
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
468 ThrValIserGlyIleAsnIleValIleGluIleuPrometIleuThrI 484
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
780 GTTATTTCAATTTCTGT.....TCATCAAG 808
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 gLyIleSerSerSpIyLeuTyrglyPheProTyrlIeAsnSerAspThr 501
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
809 CTGCATATATATGATGTTTACAGAGTTGATCTTGTAAAGATGAAGA 858
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 hrasnleuHisleuPheGluGlyIleuIle.....Lys 512
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
859 TCTCCATATACACTAAATTTTGTCTGAACAAATGACAAATATAAGA 908

```

```
513 GlyLysTyrAsn.....LeuArgAspHisThrGluAlaSerGluMe 526
    ::  |||||  ||  ::::  ||:::
909 ATTACATAGGCGACGACGATGTTATCAGCAGAACTTTTATATATTT 958
    ||:::  ::  |||||  :
526 tLeuTyrSerSer.....IleAsnTyrL 534
    :::::  ::  |||||  :
959 TTAATACAGCAAGAGAGAAATAGATCTTGTTATTCAACATACAAAT 1008
    :::::  |||  ||||:  :::::
534 euAsnSerIleLysPheLysIleAsnLysAlaMetLeuAsnPheIleLeu 550
    :::::  |||  ||||:  :::::
1009 TTGAACACTCAATCCAGATATATAGTAGATGTGACTATGCCATTT 1058
    |||  |||  |||  |||
551 AlagIuTrpAspAsnLysAsp.....SerLysLe 560
    |||  |||  |||  |||
1059 ATATTTGGTTTAAATATTAACATCACAATCATATAGGTTTTCATTAG 1108
    |::  ||  ||||:
560 uPheLysGly.....TyrAsnMetLeu..... 567
    |::  |||  ||||:
1109 AACATTAACAGAGGTGATTTAGTCCGAAATAAGAAATTATTAAGAAATTG 1158
    |||  |||  |||  |||
568 .....LysProIleLeuGluThr 573
    ||  ::|||
1159 GTAGAAACAAATACATGATTAATAATCAAAATATAAT 1197
    ::|||  |||  :::::  :::::
574 AspSerLysGluIleLysGluGluLysValSerSerAsn 586
```





```

351 AGGTGATGAGCAAAATATGATTGATTCAAACCAATGCTGCAATTT 400
|||||
117 nGlyAspArgAlaLysHisMetIleAspSerSerProMetValSerIleL 134
|||||
401 TATATATTGCATTTGCTCGCATTAATATGTTATATGATGATATGACCA 450
|||||
134 eutYrIleAlaPheAlaArgIleAsnMetLeuTyrAspValSerArgPro 150
|||||
451 TTTATGAGACAAAGATTCCTATTAGAAAACGGCTTAGAATATGA 500
|||||
151 PheAsnGlyArgGlnArgPheLeuLeuArgLysHisGlyLeuGlnTyrG 167
|||||
501 AACCTATGATGATGCTTAATGAATATGACCTATCAGAAAAGTACGC 550
|||||
167 uThrTyrGlyMetMetLeuAsnGlnIleArgArgIleArgLysValArgP 184
|||||
551 CAGATGTATATTTCTTTATCTTATAGTGGAGAAAACCTATATGATAG 600
|||||
184 roAspValIleIleLeuLeuSerLeuGlyGlnTyrTyrMetIleAsp 200
|||||
601 ATAGAAAAGAAATGATTATGTGTGATAAATATGAAGCTGTATATGA 650
|||||
201 IleGlnLysGlnIleAspTyrValAspLysIleLeuLysLeuValAsn 217
|||||
651 TTTTGATTAGATGGGTAGATATTTACTGGGACACATGGGAAGTTT 700
|||||
217 pPheAspLeuAspGlyValAspIleAspTyrGlnProHisGlyLysPhe 234
|||||
701 ACACTTAATGAATTAATTTTCAATTTATATTAATTAATTAATTAAC 750
|||||
234 YrAsnLeuAsnGlnLeuAsnPheSerAsnTyrTyrIleLysLeuIleAs 250
|||||
751 TTTGTAAGAAAACATATTCGAGAGAAAAGTTAATTTCAATTTCTGCT 800
|||||
251 LeuLeuArgLysThrIleProGlnGlnLysLeuIleSerIleSerGlySe 267
|||||
801 ATCAAATGCTCATTTATGATCGCTTTCAGAGATTCATCTTTCTGAAG 850
|||||
267 rSerAsnAlaIleLeuSerCysValSerGlyValAlaSerPheCysLys 284
|||||
851 ATGAAGATCTCCATATPACATTAATTTTGTCTGACAAATATAGAACA 900
|||||
284 spGlnGlnSerProTyrAsnThrLysPheLeuSerGlnGlnIleGlnThr 300
|||||
901 AATAAAGATTTACATAGGCGACAGCGATGTTATCAGCAGGAACCTTAT 950
|||||
301 AsnLysGlnLeuHisArgAlaAlaIleMetLeuSerAlaGlnTyrPheI 317
|||||
951 TAAATATTTTAAATACAGCAAGAGAAAATAGATCTTATTTATTTCAA 1000
|||||
317 eAsnIlePheAsnThrAlaLysGlnLysIleAspLeuValPheIleGln 334
|||||
1001 CATACAAATTTAGAACTACAAATCCAGATTAATGTTAGATATGACTTA 1050
|||||
334 hTrTyrAsnLeuGlnThrTrAsnProAspIleMetValAspMetTyrLeu 350
|||||
1051 TCCCATTTATATTTTGGTTTAAATATATACATCAATCATATTAGTTT 1100
|||||
351 SerHisLeuTyrPheGlyLeuLysTyrAsnIleThrIleIleLeuGlyP 367
|||||
1101 TTTATTTAGAACTATACAGAGGTGATTTAGTCCCGAAAATGAAGATTAT 1150
|||||
367 eSerLeuGlnHisAsnArgGlyGlyPheSerProGlnAsnLysGlnLeu 384
|||||
1151 TAGAATTTAGGAAAACAAATACATGATTAATAATCAAAATATATAGG 1200
|||||
384 euGlnLeuValGlyLysThrIleHisAspLysAsnGlnAsnAsnArg 400
|||||
1201 GCAGATGTTAGGATATGATGATTTATTTATGAAGAACATATACCAAC 1250
|||||
401 AlaAspGlyIleGlyIleThrPheHisLeuPheMetLysGlnGlnLeuPro 417

```

```

1251 TGGATCATTCGATGTAGATATTTTCTTACAAATATTTGGAAGATTTAA 1300
|||||
417 rGlySerPheAspValAspIlePheLeuThrAsnIleTyrHisIleLeu 434
|||||
1301 ATCTGGAAGTACAACTCCAAAAGACCTTACTATPACTGAAAACCTGAA 1350
|||||
434 snProGlnValGlnThrProLysAspLeuThrIleThrGlnAsnProGln 450
|||||
1351 GACTATGACACATATGATGATATGTTCCAGAGACTCGTTATTCACCAT 1400
|||||
451 AspCysSerThrIleAspGlnTyrValProGlyLeuValIleProThrI 467
|||||
1401 AGGATATATTTACAAACAAATGATGCTATATGAAAACCTAGATCTTAT 1450
|||||
467 eGlyIleTyrTyrLysHisAsnAspAlaIleTyrPyrThrArgSerTyr 484
|||||
1451 CAATTTCATGACACCTGCTGTAGACAGATATGAAATGGGACTTGTCAAG 1500
|||||
484 eTyrIleHisAlaProGlnValAspArgTyrGlnTyrPaspLeuValLys 500
|||||
1501 TGCCTATGAAAAAATATGCGATGGGAGAAAGCAGCCCATTTATTAAC 1550
|||||
501 CysTyrGlnLysIleCysAspGlyLysAlaAlaHisTyrTyrAsnThr 517
|||||
1551 CTATTAAGAAAACCTCTATTTATATATGAAAAGGGAACCATATTTAAT 1600
|||||
517 PTrLysGlnSerSerIleIleIleTyrPyrGlyGlnProTyrLeuIle 534
|||||
1601 AATGTTGGCAACAAAGACCTCCGGAAGTCCAGACATGATCTATACAA 1650
|||||
534 YStrPTrGlnGlnGlnProProGlnGlnGlnAlaLeuGlnSerTyrThr 550
|||||
1651 AAACATAGATGCATCCAAATGTCAGAGATAGAGAAATGGAATATAAATA 1700
|||||
551 LysLeuAspAlaSerLysCysProGlyTyrGlnGlnTyrPasnLysTyr 567
|||||
1701 TCCACATTAACCATAGAGTAGAGAGAACATATGAAACAGACTGATTT 1750
|||||
567 rProHisLysProLeuGlnValGlnGlnGlnTyrGlnGlnGlnValAsp 584
|||||
1751 TACCATTACAA 1761
|||||
584 euProLeuGln 587
|||||
seq_name: sp_invertebrate:090419
seq_documentation block:
ID 090419; PRELIMINARY; PRT; 378 AA.
AC 090419;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CHITINASE.
OS Plasmodium falciparum (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RX MEDLINE=20040676; PubMed=10570198;
RA Vinez J.M., Dave S.K., Specht C.A., Brameid K.A., Kuntz I.D., Xu B.,
RA Hayward R., Fidock D.A.;
RT "The chitinase PfCHT1 from the human malaria parasite Plasmodium
RT falciparum lacks proenzyme and chitin-binding domains and displays
RT unique substrate preferences."
RL Proc. Natl. Acad. Sci. U.S.A. 96:14061-14066(1999).
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC EMBL: AF172445; AAF16902.1; -.
DR InterPro: IPR001579; Chitinase.2.
DR InterPro: IPR002114; PTS_HPR_ser.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.

```





```

seq_documentation_block:
ID          P96168      PRELIMINARY;           PRT:         729 AA.
AC          P96168
DT          01-MAY-1997 (TREMBLErel_03, Created)
DT          01-MAY-1997 (TREMBLErel_03, Last sequence update)
DT          01-JUN-2001 (TREMBLErel_17, Last annotation update)
DE          CHITINASE A.
GN          ChIA.
OS          Vibrio Harveyi..
OC          Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX          NCBI_Taxid=669;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=BB7;
RX          MEDLINE=98274739; Pubmed=9611805;
RA          Swickl A.L., Kirchman D.L.;
RT          "A chitin-binding domain in a marine bacterial chitinase and other
RT          microbial chitinases: Implications for the ecology and evolution of
RT          1,4-beta-glycanases."
RL          Microbiology 144:1299-1308(1998).
DR          EMBL: U81496; AAC46383.1; -
DR          InterPro: IPR001579; Chitinase_2.
DR          InterPro: IPR001638; SBP_Dac3.
DR          InterPro: IPR003610; Chitin_bind3.
DR          PROSITE: PS01095; CHITINASE_18; UNKNOWN_1.
DR          PROSITE: PS01039; SBP_BACTERIAL_3; 1.
DR          SMART: SM00495; ChIBD3; 2.
SO          SPOUNCE       729 AA; 79016 MW; 77213CE43DB95A07 CRC64;

alignment_scores:
Quality: 163.50      Length:    610
Ratio:   0.678       Gaps:     32
Percent Similarity: 39.508      Percent Identity: 19.836

alignment_block:
US-09-579-383-2 x P96168 ..

Align seg 1/L to: P96168 from: 1 to: 729

253 AATAAAATGCTAGAAAATCCACGACAAGCAAAATTTTAGAGCAGTATAAAAA 302
||| ::::||| ::::|||
25 AsnGlySerAspMetThrAspPro.....Ty 32
303 AAGGAACAAGGTATPTAGAGCATCTACTGTCATCG..... 342
:::||||:||||:||||:||||:||||:
33 ....AspSerGlyValValValGlyTYRTrpGlnAsnTrpCysAspGlyG 48
:::||||:||||:||||:||||:||||:
343 ..AACAGTCAAGTGATGAGCAAAA.....CATATGATTGGATTCGAAC 384
::: ||||| :::||| ::::|||
48 LgYlgLYrGLngLgLYAsnAlaPrOcysValThrLeuAspGluValAsn 64
:::||||:||||:||||:||||:||||:
385 CCAATGGTGTCATTTATATATGCGATTGGCTGCAGTAAATGTTATA 434
||||| ::||| ::::||| ::::|||
65 PrometElyrAsnIleValAsnValSerPheMetLysVal.....Ty 78
:::||||:||||:||||:||||:||||:
435 TGAGTATCTAGACCATTTAATGAGA...CAAAGTTCCTATTAGAA 481
||||| ::::||| | || |||
78 rAspValAla.....AspGlyArgIleProThrPheLysLeuAsp 92
||||| ::::|||
482 AACAC...GGCTTAGATATGAAACCAATGAGTGATGCTGTTAGCAAT 528
||||| ::||| ||| ::::||||:||||:
92 roTPHrValGlyLeuSerGluGlnPhelLeuAspGlnValSerGluLeu 108
:::||||:||||:||||:||||:||||:
529 AGACGATTCAGAAAAGTAGCTCCAGATGTAAATTATCTTTATCCTTAG 578
::: ||| :::||||:||||:||||:
109 AsnLysGlnGlyArg.....SerValLeuLeuAlaLeuI 120
:::||||:||||:||||:||||:
579 TGGAGAAACCTATATGATAGATATGAA.....AAAGAAATGATTATG 622
||||| ::::||| ||| |||
120 yglYAlaAspAlaHisValGlcIleuGlnThrpGlyAspGluArgAlaPhea 137

```

```

623 TCGATAAAATATTGAAGCTTGTAAATGATTGATTTAGATGCTGATAG 672
      ::::::::::::::::::::
137 LaaspluilearLeuThrGluArgTyrGlyPheaspLysLeuasp 153
      |||||
673 ATTGACTGGGA..... 684
154 lIeaspluenglInAlaValThrAlaAaasnGlnThrValIl 170
684 ..... 684
170 eProaspAlaLeuLysLeuValLysAspHisTyrArgAlaGluGlyLys 187
685 .....CCACAT.....GGG 693
187 snPheLeuIleThrMetaIaProGluPheProTyrLeuThrGlyGly 203
      |||
694 AAG.....TTTCAACTTAAGA 713
      |||
204 LysTyrValProTyrIleAspAsnLeuGluGlyTyrTyrAspTrpIleAs 220
714 ATTAATTTTCAAT..... 729
      ::::|
220 nProGlnPheTyrAsnGluGlyLysPglYlLeTrrValaspGlyValG 237
730 .....TATTAT 735
237 lYTrpIleAlaGlnAaasnAspGluLeuLysGluGluPheIleTyr 253
736 AT.....AAATTAATTAAGTTTGA.....AAACATATCCGA 773
      |||
254 lIeSerAspSerLeuIleasnGlyThrArgGlyPheHisLysIleProH 270
774 AAAAAAGTTAAT...TCAATTTCTGTCATCAAAATGCTGCATATCAT 820
      ::::::::::::::::::::
270 snspLysLeuValPheGlyIleProSerSerIleaspAlaAla..... 284
821 GCGTTTCAGAGCTTCATCTTCTGTAAGATGAAGA..... 858
      ::::::::::::::::::::
285 .....AlaThrGlyPheValLysGluProGlnAspLeuTyrAsp 297
858 ..... 858
298 AlaPheAspSerLeuThrAlaGlnGlyInProLeuArgGlyValMetH 314
859 .....TCTCCATATA 868
314 rTrpSerIleasnTrpAspMetGlyThrAsnLysAlaGlyInAlaTyrA 331
869 ACACATAATTTTCTGTCGAACA..... 891
      |||
331 snGluGlnPheIleLysAspTyrGlyProPheValHisGlyGlnValThr 347
892 .....ATGAACAACAAATA 905
348 ProProProlValGluGlyLupPrometLeuYsgLysGluAsnThrAr 364
906 ACAATTCATATAGGCGACGAGCATGTATCAGCAGAACTTTATTAATA 955
      ::::::::::::::::::::
364 gValLeuHis.....GlyThrValPheaspP 373
956 TTTTAAAT.....ACAGCAAGAGCAAAATAGATCTT..... 987
      ::::|
373 rometGluGlyLupThrAlaThrAspLysGluAspGlyLysPheLeuThrSer 389
988 .....GTAATTAAT 995
390 SerIleAspValGluGlyTyrValGluThrSerValIleGlyThrTyrVa 406
996 TCAAAACATAC.....AATTAGAAACTACAAATCCAG 1027
      ::::::::::::::::::::
406 lIeuThrTyrArgValLysAspSerAspAsnAsnGluThrThrLysAlaA 423
1028 ATATAATGCTAGATATGTAATCCCATTTATATTGGTTAAATAT 1077

```

```

423 rg...ThValGluValTyrSerGlnLysProValPheaspGlyValSer 438
      |||
1078 AACATCACAAATCATATATAGGTTTTCATTAGAACATAACAGAGGAGATT 1127
      ::::::::::::::::::::
439 AsprThrThrValValLeuGlyAsnSerPheasp...PrometAlaGlyVa 454
1128 TAGTCCCGAAAAATAAATAATTAATTAATGTAATGTTAGAAAAACATATCATG 1177
      ::::::::::::::::::::
454 lThrAlaAsnAspAlaGluAspGlyAspLeuThrSerSerIleIleHis. 470
1178 ATAAAAATCAAAATATATATAGGCGAGATGCTATAGGATATGCAATTTA 1227
470 ..... 470
1228 TTTATGAAGAACAATATCAACATGATCATTCGATGAT..... 1269
      |||||
471 .....ThrGlySerValAspValAsnGluIleGl 480
1270 .....ATTTCCTTACAAATATTTGAAACATTTAAATCCTGAAGTAC 1312
      ::::|
480 YAsnTyrThrLeuValTyrArgValThrAspSerAlaAsnGlnThrValT 497
1313 AAATCCAAAAGACCTTACTATATGTAAGAAACCTGAA..... 1350
      ::::|
497 hrAlaGluArgLysValThrValThrAspGlySerAsnCysAlaAlaAla 513
1351 ...GACGTAGACAAATAGATGATATGTTCCAGGACCTCGTATATCCAAC 1397
      |||||
514 TrpAspAlaAsnThrVal.....TyrValGluGlyAspGlnVal..... 526
1398 CATAGGATATATATTAACAACAACATGATCTATATGAAACTAGATCTT 1447
      ::::::::::::::::::::
527 .....SerHisAspGlyAlaThrTrpValAlaGlyTrp 538
1448 ATTCAT...CATGCACCTGCTGTAGACAGATATGATGGACATGCTC 1494
      ::::::::::::::::::::
538 yTrhArgGlyGluLupProGlyThrThrGly...GluTrpGlyValTrp 553
1495 AAAGTATGCTATGAAAAATATATGCGATGG 1524
      |||
554 LysLysAlaSerAspSerSerCysGlyGly 563
seq_name: sp_organelle:034192
seq_documentation_block:
ID 034192 PRELIMINARY; PRT; 590 AA.
AC 034192:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 5.
GN ND5.
OS Crithidia oncopelti.
OG Mitochondrion.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX NCBI_TaxID=5657;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-068;
RA Dmitriy Maslov A.;
RL Submitted (OCT-1990) to the EMBL/GenBank/DBJ databases.
RP [2]
RC STRAIN=S-068;
RA Maslov D.A.; Horvath A.; Gwang II K.; Kolesnikov A.A.;
RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.
CC -i- CATALYTIC ACTIVITY: NADH + UBIOQUINONE = NAD(+) + UBIOQUINOL.
CC -i- SIMILARITY: TO NADH-UBIOQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS CHAINS.
EMBL: X56015: CAA39492.1: -.
DR InterPro: IPR001750: Oxidored_q1.
DR Pfam: PF00361: oxidored_q1; 1.

```

KW Mitochondrion: NAD: Oxidoreductase: Ubiquinone  
 SO SEQUENCE 590 AA; 71650 MM; 1A2FE619E5B0C02E CAC64;

## alignment\_scores:

Quality: 158.50 Length: 681  
 Ratio: 0.576 Gaps: 41  
 Percent Similarity: 40.382 Percent Identity: 21.145

## alignment\_block:

us-09-579-383-2/rev x Q34192 ..

Align seg 1/1 to: Q34192 from: 1 to: 590

```

1743 TTCTTGTCATATGTCTCTACTTCTAGTGTTATGATATTTT 1694
      |||:|||||:|||||:|||||:|||||:|||||:|||||
4 phephepheleuLeuPheLeuPheGlyPhePheLeuGlyThrPhe 20
1693 T.....ATTCATCTCTCATGCC..... 1675
      |||:|||||:|||||:|||||:|||||:|||||:|||||
20 eleuGlyArgHisLeuLeuSerPheThrLeuSerIleLeuMetThrValP 37
1674 .....TGACATTTGGATGCATCATGTTTGTGTA 1645
      |||:|||||:|||||:|||||:|||||:|||||:|||||
37 heLeuValMetValThrMetPheSerPhePheCysIleSerValCysLeu 53
1644 TGACTCTATGTC...CTGACCTCCGAGAGCTCTTGCCACCATTTAA 1598
      |||:|||||:|||||:|||||:|||||:|||||:|||||
54 TyrGlyTyrCysTyrTyrAspPheCysLeuIleLeuMetLeuAspLeu 70
1597 TTAATATAGTCCCTTCCATATATATAGTCTTCTTATAGTCA 1548
      |||:|||||:|||||:|||||:|||||:|||||:|||||
70 sPheIlePheMetSerPheTyrCysAsnGlyPheTyrLeuPheIleLeuP 87
1547 GTGTATA...ATAATGGCTGCTTCCATCGCATAT...TTTTCATA 1504
      |||:|||||:|||||:|||||:|||||:|||||:|||||
87 heLeuIleAspLeuValPheCysPheIleLeuPheTyrAlaPheTyrTyr 103
1503 GCATACCTTGACCAAGTC.....CCATTCATATCTGTC.....TA 1469
      |||:|||||:|||||:|||||:|||||:|||||:|||||
104 MetTyrTyrAspLeuMetLeuLysArgPhePheAsnIlePheTrpTrp 120
1468 CACGAGGTCATGATTAATAGATCTAGTTTCCATATAGCATCATG 1419
      |||:|||||:|||||:|||||:|||||:|||||:|||||
120 eValLeuCysMetAsnPhePheIleLeuSerTyrAspTyr..... 133
1418 TGTTTGAATATATCCCTAT...GGTTGGAATACGAGCTCTGGAACATA 1372
      |||:|||||:|||||:|||||:|||||:|||||:|||||
134 .....LeuThrAlaTyrCysGlyTrp.....GluLeuLeuGlyLeu 145
1371 TTCACTATGTGCTACAGTCTTCAAGGTTTTCAGTTATAGTAGGCTT 1322
      |||:|||||:|||||:|||||:|||||:|||||:|||||
146 PheSerPhe.....PheLeuIleSerTyrPheTrpTyrArgP 158
1321 TTGGAGTTTGTACTTCAGATTTAATGTTCACAAATTTGTAAGAAA 1272
      |||:|||||:|||||:|||||:|||||:|||||:|||||
158 ePheAlaLeuLysPhe.GlyPheLysSerPhe.....PheIleSerLys 172
1271 ATATCTACATGCAAGATCCAGTTGTAATGTTCTTTCATTAATAATG 1222
      |||:|||||:|||||:|||||:|||||:|||||:|||||
173 IleGlyAspVal..LeuLeuLeuLeuSerPheValMetThrPheIleSer 189
1221 CC.....ATATCCCATATACCATCTGCCCTATATATATTTGAT 1184
      |||:|||||:|||||:|||||:|||||:|||||:|||||
189 hrGlyTyrGlyMetIleAsnPheTyrPheValAsnPheLeuCysValAsp 205
1183 TTTTATCATGATGTTTTCCTACCAATTAATTAATTTCTTAATTTG 1135
      |||:|||||:|||||:|||||:|||||:|||||:|||||
206 PheTyrPheThrAlaPheMetLeuPheLeuLeuIleMetCysAlaPheThr 222
1134 .....GGACTAAATCCACCTCTGTATGTTCTAATGAAA 1100
      |||:|||||:|||||:|||||:|||||:|||||:|||||
222 rLysSerThrGlnPheGlyLeuHisIleTrpLeuProAspAlaMetGluG 239

```

```

1099 AACCT.....AATATGATGTG... 1083
      |||:|||||:|||||:|||||:|||||:|||||:|||||
239 lYProIleProValSerAlaLeuIleHisAlaIaThrLeuValValCys 255
1082 .....ATGCT 1078
256 GlyIleLeuLeuValSerPhePhePheTrpCysPheAspPheTrpLeuAl 272
1077 AATTTTAAACCAAAATATAATGGATAGTACATATCTACCATTAAT 1028
      |||:|||||:|||||:|||||:|||||:|||||:|||||
272 eTyrPheTyrProLeuIleGlyTrpSerSerLeuIleLeuValMetMetS 289
1027 CTGATTTGT.....AGTTCT 1011
      |||:|||||:|||||:|||||:|||||:|||||:|||||
289 er..LeuCysValPheTyrAsnPheAspAlaLysArgPheValAlaPheS 305
1010 AATTTGTATGTTTGAATAATAACAAGATCTTTTCCTCTTG... 967
      |||:|||||:|||||:|||||:|||||:|||||:|||||
305 eThrIleCysGlnIleSerPheSerMetPheCysCysLeuCysLeuAsp 321
966 .....TGATTAATAAATATTAATAAAGTTCTGCTGATACA 929
      |||:|||||:|||||:|||||:|||||:|||||:|||||
322 LeuTyrValGlyCysLeu..... 327
928 TCGCTGCTGCCCTATGTAATCTTATTTGTTCTATTTGTCAGACAAA 879
      |||:|||||:|||||:|||||:|||||:|||||:|||||
328 .....PhePheCysTyrHisMetPheTyrLysA 337
878 AA.....TTTAGTGT.....ATATGG..... 862
      |||:|||||:|||||:|||||:|||||:|||||:|||||
337 lathrLeuPheIleValIleuGlyValTrpIleHisLeuPheGlyLeu 353
861 .....AGATCTTCATCTTTCACAGAAAGATGC..... 835
354 GlnAspValArgCysTyrPhePheIleTyrPheCysGlyCysValIleuAl 370
834 .....ACCTCTGAAACGATGATTAAT 813
370 aArgMetLeuLeuValPheAlaLeuLeuAsnSer.....C 382
812 GCAGATTTGATGAACCAAGAAATTAATTAATCTTCTCCGGAATAGT 763
      |||:|||||:|||||:|||||:|||||:|||||:|||||
382 ySer..... 383
762 TTTTCTTAACAAGTTAATTAATTAATAATAATTTGAAAAATTTAAT 713
      |||:|||||:|||||:|||||:|||||:|||||:|||||
383 ..... 383
712 CATTTAAGTTAAAACTCCCATGCTGTTCCAGTCAATATCTACCA 663
      |||:|||||:|||||:|||||:|||||:|||||:|||||
384 .....LeuTrpPheLeuCysGlyPheTyrCysL 393
662 TCTAATCAAAATCATTAACAAGCTTCATATTTTATCCAC...ATPAC 616
      |||:|||||:|||||:|||||:|||||:|||||:|||||
393 yAspLeuLeuLeuCysThrLeuMetLeuValSerPheHisPheIleLeu 409
615 AATTTCTTTTCTAT...ATCTATCATATAGGTTTCTCCACCTAAGATA 569
      |||:|||||:|||||:|||||:|||||:|||||:|||||
410 GluPheLeuPheValCysIlePhePheIlePhePheThrAla..... 423
568 AAAGAAATTAATTAATCTGAGACGTATTTCTGATACGCTCAATTTATTA 519
      |||:|||||:|||||:|||||:|||||:|||||:|||||
424 .....IleTyrAsnTyr..PheLeuLeuPhePheLeu..... 433
518 AGCATCATATACATAGTTTCATATNTCAATACCGGTTTCTTATATAGAA 469
      |||:|||||:|||||:|||||:|||||:|||||:|||||
434 .....CysPheValAlaPheLysCys 439
468 TCTTTGCTTCATTAATATGCTAGATACATCATATATACATATTAATGC 419
      |||:|||||:|||||:|||||:|||||:|||||:|||||
439 sPheCysLeu..... 442

```

```

418 GAGCAATGCATATATAAATTACACCATTTGGCTTGAATCA..... 375
443 ..AlaaspCysleuPheleu.....PheaspPheGluCysCysleu 456
374 ATCATGTTTGGCTGCTATCATCCTTGACTGATGATACCATGATGCC 325
457 IleTyrcysPhe..LeuGlyleuTyrcys.....PheIlele 469
324 TGCATATAACCTGTTCCCTTTTATATCTCCTAAATTTGCTGTG 275
469 euIlePhePheValIleaspPheleuTyrllePheValPheSerSerTyr 485
274 GTGATTTTCTACATTTTATTTATTTATTTATTTTGTAAATCTA 225
486 CysMetPheTyrSerPheTyrTyrTyrPheaspPheaspIleAl 502
224 GAGTTACACCATCTCCGCAATAAGATCCATTTCTACAAATTTACAT 175
502 aserPheThr..... 505
174 ATTCGATTTAAGATGTGMAAAGACATCATTTCCGTTTGATGACTTT 125
506 .....ValPheValMetIleSerleuAlaPhe 514
124 TATTTTATTTTCCCGTATTTATTCACATGATTTATTTATTTT 75
515 IleTyrlTyrGlyCysValleuPhe.....TyrPhePhe 525
74 C.....CTTCAAGGTCTCGAATTTG 53
525 easnValaspCysIleMetleuPheThrPargIle 536

```

seq\_name: sp\_invertebrate:097230

```

seq_documentation_block:
ID 097230 PRELIMINARY: PRF: 1410 AA.
AC 097230:
DT 01-MAY-1999 (TREMBLER, 10, Created)
DT 01-OCT-2000 (TREMBLER, 15, Last sequence update)
DE 01-JUN-2001 (TREMBLER, 17, Last annotation update)
OC HYPOTHELICAL PROTEIN, PFC0195W.
OC Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jaseal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skellton J., Squares R., Squares S., Sultson J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.:
RT *The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.
RT Nature 400:532-538(1999).
RL EMBL, AL034558; CAB38995.2; -.
KW Hypothetical protein.
SQ SEQUENCE 1410 AA; 170544 MW; A1590A8CB8011EB3 CRC64;

```

alignment\_scores:

Quality:	154.00	Length:	651
Ratio:	0.526	Gaps:	30
Percent Similarity:	45.008	Percent Identity:	20.123

alignment\_block:

US-09-579-383-2 x 097230 ..

Align seg 1/1 to: 097230 from: 1 to: 1410

```

31 ATGATTCATCTTTGTATTTCTGCAAAATTCAGAACCTTGAAAGAAAAA 80
111 .....
86 IleGlnIleIleArgHisSerAsnIleYaspAsnSerSerAsnAsn 102
81 TAATATTAATAATTCATTTGGGAATAATACGGAAATAATAATAAATC 130
102 dnAsnAsnAsn.....AsnAsnSerAsn 112
131 ATCAAAAGGAATAATAGATGCTTTTTCACATCTTAATGAATAATAGT 180
112 snAsnAsnAsnAsnAsnSerAsnAsnAsnAsnAsnSerAsnAsn 128
181 AATTTGTAGAAATGAGTCTTATTCGGAGATGGGTACTCTGAAAT 230
129 Asn.....AsnSerAsn 133
231 TACAAAAAATAATAATAATAATAATAATAATAATAATAATAATAA 280
133 dnAsnAsnAsnSerAsnIleYaspGlyIleAsnIleValIleAsnThr 162
281 AATTTTACAGAGATATATAATAATAATAATAATAATAATAATAATA 324
148 .....TyrAsnAspTyrTyrIleYaspGlyIleAsnIleValIleAsnThr 162
325 GGAATCTAGTTCATGGAACAGTCAAGGTATAGACAAATAATATGAT 374
163 AsnTyrThrSerSerTyrTyrIleYaspGlyIleYaspGlyIle 179
375 TGATTCAAACCAATGCTGATTTTATATATGATTTGTCGCGCATTA 424
179 eIleAsnAsnAspValPheSerSerLeuGlnAsnIlePheleuIle 196
425 ATATGTTA.....TATGATGTATCTAGACCATTTAATGA 459
196 snIleAsnIleYaspPheGlnArgValGlnProIleYaspGly 212
460 AGACAAAGATTCCTATTAAGAAAAACACGCTTGAAATATAAACCTAT 509
213 AsnGlnIleGlnIle..... 218
510 TATGATCTTAATGAATTAATGACGTATCAGAAAGATGCTGAGATTA 559
219 .....IleAspAsnValIleYaspGlyIleAsnIleValIle 232
560 TTAATCTTTTATCCTTATGATGAGAAACCTATGATATATAGAAAA 609
232 euHisIleleu.....LeuYaspAsnAsnValAsnIleAsnIleTyrIle 247
610 GAATGTATATGCTG.....GATAAATATGTAAGCTTGTATATGA 630
248 GluIleAspIleleuIleuAlaSerGlnTyrPheSerHisLeuCysAsn 263
651 TTTTGAATTTGATGCTGATATTTGACTGGAACCATGGAAGTTT 700
263 ..... 263
701 ACAATTAATGAATTAATTTTCAATTTATTAATTAATTAATTAAC 750
264 ..AsnIleAsnAsnLeuAspPhe.....PheSerleuValSer 275
751 TTGTTAAGAAAACTATTCGGAAGAAAGTTA.....ATTTCATTTTC 794
276 LeuPheHisIleTyrIleProIleYaspGlyIleYaspGlyIleVal 292
795 TGCTTCATCAATGCTGATTTATGATGCTGTTAGAGAGTTGATCTTTC 844
292 IleYaspAsnAsnAspIleTyrAsnSerAsnAsnSerTyrAspGlyIle 309
845 GTAAA.....GATGAAGAAATCTCCATATAACACTAAA 876
309 IleYaspAsnAspGlnIleAsnIleAspGlnIleGlnIleYaspGlnIle 325
877 TTTTGTCT..... 885

```

```

326 HistLeuAsnLeuIleProPheLeuAspIleIleLysAsnCysIleLeuLys 342
886 .....GAAACAATAGACAACAATA 904
342 sAsnIleCysIleGlnIleAsnHisLeuMetThrAsnValLysAsnAsnA 359
905 AAGAATTTCATAGGCGACAGCGATGTTATGACAGAACTTTTATTTAT 954
359 sngIu.....HisLysAsp.....Asn 364
955 ATTTTATATACGACGAAGAATAAGATCTGTATTTATTCACAACATA 1004
365 ProIleAsnValAsnLysLysLysLysValThrPheHisAspIleTy 381
1005 C.....AATTGAAACTACAAATCCAGATATTAATGCTAGATA 1042
381 rGlyGluValIleAsnIleLeuTyrlleIleLysAspAspIleLysA 398
1043 TGTACTTATCCATTATATTGTTTA..... 1071
398 sntYrIleAsnTyrlleHisValIleLeuAspThrLeuTyrlLysAsnIle 414
1072 .....AAATATAC..... 1080
415 HisLysHisIleHisIleHisTyrlAsnTyrlThrTyrlSerIleLysLeuAr 431
1081 ....ATGCATATCATATTAGGTTTTCATAGAACATAACAGAGTGGAT 1126
431 gTyrlleAsnThrIleLeuSerPheAsnIleAsnHisSer..... 444
1127 TTAGTCCCGAAATAAAGATATTATAGATTGTTAGGAAACAATACAT 1176
445 .....TyrlleSerAsnGlyLysIleArg 453
1177 GATTAATAATCAATAATATATAGGCGAGATGTATAGGATATGGAATT 1226
454 GluGlnIuLysAsnValSerArgTyrlPheSerIleIleAspAsnAs 470
1227 ATTTAGAAAGAACATATACCACTGATTCATTCGATGTA..... 1266
470 nPheLeuTyrlleArgLysGluAsnPheSerPheAsnTyrlValHisLysA 487
1267 .....GATATTTT...CTTACAATATTTGG 1290
487 sntYrAsnIleLeuTyrlAsnThrPheMetTyrlAsnThrIleAsnIleAsn 503
1291 AACATTTAATCTCTGAA.....GTACAACCTCCAAAGACCTTAC 1331
504 AsnAsnIleAsnHisAspLysAsnIleGlnGlnTyrlThrAsnAspAs 520
1332 TATTAAGTGAACCTGAGACTGTAGCACAATGATGATAT.....G 1375
520 nThrIleTyrlleGluAspLeuAsnLeuGlnAsnTyrlAsnLeuA 537
1376 TTCAGAGACTCGTTATTCACAC.....ATAGGGATA 1407
537 sntLysLysLeuLeuTyrlLysThrAsnGluAspIleTyrlAspIleSerPhe 553
1408 TATTACAACAC.....AATGATGC 1427
554 MetProLysHisIleGlnHisAsnAsnIleLeuAsnAspAsnAspVa 570
1428 TATATGAAAACTAGATCTTATTC...ATTGATGCACTGTGTAGACA 1474
570 lThrTrpAsnAsnSerSerSerAsnTyrlHisIleProAsnValTyrl 587
1475 GATATGAATGGAC..... 1488
587 ysgLnuIuTyrlAspLysMetGlyThrLeuProIleIleLysAsnSerPro 603
1488 ..... 1488

```

```

604 AsnGluTyrlAsnLeuIleSerLysAspIleGluSerTyrlLysTyrlleI 620
1489 .....TTGTCGAAGATGCTATGAAGAAAATATATGCGATG 1522
620 elysPheSerIleGluLeuLeuTyrlleIleValLysLysIle..... 634
1523 GGAAGACAGCCCATTTATTAACACGTAGCTATAAAGAACTTATTT 1572
635 ..LysTyrlValHisIleHisThrAsnThr...PheAlaAspGluIleIle 649
1573 ATA 1575
650 Leu 650

```

seq\_name: sp\_Invertebrate:097291

seq\_documentation\_block:

```

ID 097291 PRELIMINARY; PRT; 1946 AA.
AC 097291;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 231.8 KDA PROTEIN.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagsels K., Jassal B., Kyes S., Mclean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sultson J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538 (1999).
DR EMBL; AL034559; CAB39037.2; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Hypothetical Protein.
SQ SEQUENCE 1946 AA; 231792 MW; 59AC248AB9808B34 CRC64;

```

alignment\_scores:

```

Quality: 150.50 Length: 629
Ratio: 0.510 Gaps: 26
Percent Similarity: 46.900 Percent Identity: 19.873

```

alignment\_block:

US-09-579-383-2 x 097291 ..

Align seg 1/1 to: 097291 from: 1 to: 1946

```

46 TATTTCGAATATTCGAAACCTGAAAGGAAAAATATATATGATTC 95
1030 TyrlArgAspAsnAsnArg.....AsnLysAsnAsnTyrlArgAspAs 1043
96 ATTTGGGAATATATGCGGAAATATTAATAAATCAATCAACGGAATATC 145
1043 n.....AsnArgAsnLysAsnAsnTyrlArgAspLysA 1054
146 ATAGAGCTTTTTCACATCTTAATTCGAATATATGATAT...TTTGTAGAA 192
1054 snHisAsnSerAsnAsnAsnLysAsnAsnAsnLysAsnLysAsnAsn 1070
193 TATGAGATCTTATTCGCGAGATGGCTAATCTGACGATATACA..... 234
1071 TyrlTyrlTyrlGlnHisAsnAsnAsnLeuSerHisIleThrValLeuGln 1087

```

[illegible]

```

1339 uasnVllysLeuIle.....GluThrAsnLysThrLeuT 1351
1127 TTATGCCGCAAAATAAGAAATTATAGAAATGGTAGGAAAAACAATACAT 1176
1351 yrlle...GluAsnGluLysLeuSerAsnAspLeuLysAsnThrLeuLys 1366
1177 GATNAAAATCAAAATATATATATAGGGCAGATGCTATAGGAGATATGCGATT 1226
1367 GluLysGluLysLleGlnGluAsnTyLysAsnLysLleAsnLysGluHisL 1383
1227 ATTATGAAAGAACAAATTACCACTGATCATTCGATGATGATATTTC 1276
1383 elysLlelleglGluLeuGlnSer..... 1391
1277 TTACAAATATTTGGAAACATTTAATCTGAAGTACCAAACTCCAAAGAC 1326
1392 .....TyrLysGluLysGluLysHisMet 1400
1327 CTATCATATACATAAAACCTGGAAGACTGTAGCAGCAATATGATATGCT 1376
1401 LysGlnlIleThrLleLysGluGlnLleHisIsnLeuAspLysHisL 1417
1377 TCCAGACCTGCTATTCCCAACCATAGAGATATATTACAAACACATGATG 1426
1417 ealaglnlleglnlleglu.....LysAsnAsn.... 1426
1427 CTATATGAAAAACATAGATCTTATCAATTCATGACACTGGTGTAGACAGA 1476
1426 ..... 1426
1477 TATGATGGGACTTGGTGCACAAATGATGCTATGAAAAAATATGCCATGGGAA 1526
1427 PheglGluSerTyrLeuLys.....GluLysAsnGluAsnGluL 1440
1527 AGCAGCCCATATTATTAACCTGACTGATTAAGAAAGCTCT..... 1566
1440 smetSerAsnLleLeuGlnGluLysTyrLysGluLeuSerThrTyrGluI 1457
1566 ..... 1566
1457 leAspLysAsnLleAsnLysLleLysLleGluAspLeuGluLysAspLys 1473
1567 .....ATTATTTATGGAAGGGAACCATAT.....TT 1555
1474 GluAsnLleLeuLeuThrLysAsnGluGluLleAsnAsnLeuLysGluI 1490
1596 AATTAAATGGTGCAACCAAGACCTCCGGAAGGTCAAGCACTAGACAT 1645
1490 uTyTyLysMetValGlnGlnHisLeuGluAspThrAsnValLeuTyrGluL 1507
1646 ACACAAACTAGAT.....GCA 1662
1507 yslInLysLeuAlaLleAspThrLleThrLysGluLysAsnAsnLleLle 1523
1663 TCCAAATGTCACGAGTAGAAGATGGAATAAATAATCCACATTAACC 1712
1524 AsnGluCysAspLysLleLysAsnLysAsnLysLysLeuAsnAsnLysLe 1540
1713 ACTAGAAGTAGAGGAACAATATGACACAGAAGTCGAT 1749
1540 uLysGluAsnGlnAsnAsnTyrGlnHisThrLeuAsn 1552
seq_name: sp_invertebrate:077393

seq_documentation_block:
ID 077393; PRELIMINARY; PRT: 2423 AA.
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHEtical 293.7 kDa PROTEIN.
OS Plasmodium falciparum (isolate 3D7).

```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
 RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagsall K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy J., Oliver K., Quail M.A., Rajandream M.-A.,  
 RA Ruter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 falciparum".  
 RL Nature 400:532-538(1999).  
 DR EMBL: Z98551; CAB1149.2; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 2423 AA; 293711 MW; 0FED1B18FC27A2F6 CRC64;

alignment\_scores:  
 Quality: 150.50 Length: 655  
 Ratio: 0.514 Gaps: 38  
 Percent Similarity: 44.733 Percent Identity: 19.847

alignment\_block:  
 US-09-579-383-2 x 077393 ..

Align seg 1/1 to: 077393 from: 1 to: 2423

```

76 AAAAATAATATATATATTCATTGGGAAAT..ATACGGAAATATAAAA 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1542 LysAsnIleValAsnIlePylsIleTyrGluValIleArgIleGluLysAs 1558
   |||  :|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 TAAAACTCATCAACGAAATATACATGCTTTTTCATCTTAATCGA 172
   |||  :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1558 nIySlySAsnIySAsnArgIleAsnIyS..TyrSerTyr.....A 1571
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
173 ATAAATAGTAATTTTGTAGAATATGATCTTATTCGCGAGAT..... 216
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1571 snAsnIySgluTyrIle.....LeuIySAsPleuArgArg 1582
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 TGTAACTCTAGAAATTCACAAAATAATATAA..... 246
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1583 CysGlnIySlySleIySlySAsnAsnIySleIleIleIySlySleAsnAs 1599
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 .....AATATAATATAAATGATAGAAAATACACCAAGAC 280
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1599 nIyScysAsnMetTyrAsnMetAsnAsnIleTyrSerArgIleLeuArgA 1616
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 AAATTTTACAGAGATATAAAAAAGAAACAGGATTTATACAGAGATAC 330
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1616 snPheLeuAsnIySAsnIySlySArgIySAsnIySleuPheSerSerTyr 1632
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 TATGTTTCATGGAACGTCACAGTGTATAGCA.....AACATATGAT 374
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1633 ValAsnIleLeuAsnAspCysLeuHisArgTnIleArgIyShIleIle 1649
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
375 TGATTCAAACCAATGGTGTCAATTTATATATGATTCATTGTCGCGATTA 424
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1649 elySAsnAsn.....PheIle..... 1654
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 ATATGTTATATGATGATCTAGACCAATTTATGAGAGACAAAGATTCTTA 474
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1655 .....TyrAsnAsnIySgluAsn..... 1660
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
475 TTAAGAAAAACGCGCTTAGAATATGAACCTATGATATAGCTTAATAGA 524
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1660 ..... 1660
525 AATTAGAGTATCAGAAAAAGTAGTCGACGATGTAATTTATCTTTTATCT 574
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

1661 .....AsnIySleValGlnAsPleIySgluLeuTyrSerL 1673
575 TAGGTGGAGAAACCTATATGATATATAGAA..... 606
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1673 euSerSerValHisAsnIleArgAsPleIleGluCysGluTyrGlnGluArg 1689
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
607 .....AAGAAATTGATTTATGTGATTAATAATATTCG.....AA 638
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1690 MetArgIySgluMetLeuIySlyPheAsPleIleIleAsPHisIySgluIy 1706
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
639 GCTTGTTAATGATTTTCAT.....TTAGATGCTGATGATATTTGAC. 678
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1706 sleuIleAsnSerPheAsnIyValCysValAsPAsnIleSerLeuGlyL 1723
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
678 ..... 678
1723 euHisAsnAsnAsnAsnIySlyAlaLeuGluAsnIySgluIySser 1739
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
679 .....TGGAGACACATGGGAGATTTCACAACTTAATTAATTAATTT 722
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1740 GluAlaTyrIySProTyrGlySerAsnIySAsnIleAsn..... 1752
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
723 TTCAAATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTCG 772
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1753 .....LysAsnIleAsn.....IleAsnIySAsnIleAsnI 1763
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
773 AAGAAAGTTAATTTCAATTTCTGTTCATCAATGCTGCATTAATCAATGC 822
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1763 leAsnIySAsnIleAsnIleAsnIleAsnIleAsnIleAsnIySAsnIy 1779
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
823 GTTTCAGAGCTGCATCTTTCTGTAAGATG.....AAGATCTGCC 863
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1779 IAsnIleAsnIySAsnValAsnIleAsnIleAsnIleAsnIleAsnIle 1796
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
864 ATATAACACATAATTTTGTCTGCAACAATAGAAACAATTAAGATTAAC 913
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1796 snIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAs 1812
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
914 ATAGGCGACGACGATTTATCAGCAGAACTTT..... 948
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1813 CysAsnHisGluAsnIySAsnArgGluAsPleIleHisTyrAsnHisIsty 1829
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
949 .....ATTAAATTTTATATACAGCAAGAGAAATAGATCTTG 988
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1829 rIyShIsmetIleTyrPheHisAsnIySlySAsnAsPlySerGluLeuL 1846
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
989 TA..... 990
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1846 euIySlySAsnSerGlyGluLeuGluTyrAsnHisGluGluArgIySlyr 1862
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
990 ..... 990
1863 LysGluIySleIyTyrSerCysIySArgAsnPheGluIleGluAsnIleAs 1879
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
991 .....TTTATTCAA.....ACATACAAATTAGAAA 1015
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1879 nTyrPheIleTyrArgAsnIleTyrArgProIleIySlyrHisIyArgP 1896
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1016 CTACAAATCCAGATATTAATGATAGATATGATCTTATTCACAT..... 1056
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1896 IeThrAsn...LysIleLeuIleAsnSerPheLeuIleAsnMetAsnLeu 1911
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1057 .....TTATATTTTGTCTTAATAATATACATCAACATCATATTAAGTTTTC 1103
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1912 LeuIleTyrPhe.....TyrIleLeuIyScysIleLeuPheAsPAs 1925
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1104 ATTAGAACATAACAGAGTGCA.....TTTAGTCCCGAAA 1138
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1925 pMetAsPAsPAsnTyrAsnPheLeuCysAsnIleLeuPheAsnIySArgA 1942
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1139 ATTAAGAAATTAATTAATTTGTAGAGAAAAACATACATGATTAATAATCA 1188
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1942 snIyS.....IleIleAsnIyS.....IleGluAsnAsPAsPlyr 1953

```



```

1189 AATAATAATAGGCGAGATGATAGGATATGCGATTTATTTATGAAAGA 1238
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1194 AsnaSpasncysCysasp.....HisilephevalAsnly 1965
1239 ACAATTACCACTGGATTCATTCATGATAGTATATTTCTTCTCA...AATA 1285
      :.....:|||||:|||||:|||||:|||||:|||||:|||||:
1965 s.....:AsnlySHismetvalGlnlyrPheleuAsnlyAsnp 1978
1286 TTTGGAAACATTTAAATCCTGAGTACAAACCTCAAAAGACCTTACTATA 1335
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1978 hetYrThSerCysasn.....IleYsAsnlyscysval 1989
1336 ACTGAAACCCCT.....GAAAGCTGTACACATGATGATGATATGT 1376
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1990 AsplysAsnlyleuTYrTYrThraspIleSerThrIle..... 2002
1377 TCACAGACTCGTTATTCACACCATGSGATATATATACAAACATGATG 1426
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2003 .....PhehetYrIleHis..... 2007
1427 CTATATGGAACATAGATCTTATTCATTCATTCAGCACTGTGTAGACAGA 1476
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2008 .....PheYsIleAsnValPro..... 2013
1477 TATGATGGGACTTGCTCAAGATATGCTATGAAAAAATATGCGAGGAA 1526
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2014 .....LysleuLeuArgIlePhephe.....AspHisly 2023
1527 AGCAGCCATATTTATACACTGACTATAAAGAAAGCTCTATAT... 1572
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2023 sValValAsnlyTYrAsnAspIleTYrArgIleuThSerleuLeuTYrP 2040
1573 .....ATATGG 1578
2040 heserleuLeuAsnleuAsnTYrIlePheleuIlePheSerThleucys 2056
1579 AAAGGGGACCATATTTAATTAATGTGTGCAACAGAGCACTCCG.... 1623
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2057 ArgProThrIlePheGlnCysIleYsIleStrpYsGlyAsnProSerTh 2073
1624 ...GAAGTCAG 1632
2073 rAlaGlnIleGln 2077
seq_name: sp_fung1:006350
seq_documentation block:
ID 006350 PRELIMINARY; PRT; 511 AA.
AC 006350;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO CHITINASE.
GN YDR371W OR D9481.7.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Ding H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis S., Meneses S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Telch A., Trevasakis E., Vignati D., Wilcox L., Mohlman P., Vaudin M.,

```

```

RA Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Jia Y., Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: U28373; AAB64807.1; -.
DR SGD: S0002779; YDR371W.
DR InterPro: IPR001579; Chitinase.2.
DR InterPro: IPR001223; Glyco_hydro_18.
DR Pfam: PF00704; Glyco_hydro_18; 2.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR GlycoSite: HydroLase.
SQ
SEQUENCE 511 AA; 5918 MW; 238PF7963BFD63 CRC64;

alignment_scores:
Quality: 149.00 Length: 599
Ratio: 0.598 Gaps: 28
Percent Similarity: 41.569 Percent Identity: 18.698

alignment_block:
US-09-579-383-2 x 006350 ..
Align seg 1/1 to: 006350 from: 1 to: 511

140 AATACATGAGCTCTTTTCACATCTTAATGAAATTAATTTTGTGA 189
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
7 GlnHisArgSerLysSerSerleuValSerHisleuIleleuIle 23
190 GAATATGATCTTATTTGCGAGATGGGTGT...AATCTAAGATTAACAA 236
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
23 ePheIleThrIleIleIleGlnMet..CysleuTYrAsnLysIlePhe 39
237 AATATTAATAA.....AATATAATAATAAATGATAGAA 268
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
39 sAsnGlnArgSerAspAspIleArgAspAsnPhaAsnAsnGlyGlnHis 56
269 AATCACCAGACAAATTTAGGAGGATATAAAAA...AGGAACAAGGT 315
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
56 rGValProSerAsnValGlnAsnHisGlyThrHisIleArgAspGlnAla 72
316 ATTTATGACAGA...TACTATGTTTCATGGAACAGTCAGGTGATAGAGC 362
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
73 PheIleSerGlyValTYrTYrSerAsnTrpSerProTYrLysProArgPh 89
363 AAAACATATGATGATTCGAAC...CCAAATGGTCAATTTTATATATG 409
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
89 eHisPheProHisAspIleAsnleuLysGlnValSerHisIleTYrTYr 106
410 CATTTGCTGCATTAATATGATGATATGATGATAGCCATTTAATGGA 459
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
106 lApePheLysIleAsnSer..... 112
460 AGACAAGATTCCTATTAGAAACACGGCTTAGAA..... 495
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
113 .....ArgThGlyGlyIleGlnAsnThrAspSerTr 123
496 .....TATGAACCTATGATGATGCTTATAG 523
123 pSerAspLeuGlnMetAsnleuTYrLysSerleuAlaIleLysAsnSerG 140
524 AATAT.....AGACGTATCAGAAAAGTACGCCA... 552
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
140 luleuIleLysGlnSerSerAsnAsnSerValGlnAsnIleleuProleu 156

```

```

553 .....GA 554
157 GlycylleuLeuPheTyrLeuLysThrCysSerAspLysLeu 173
555 TGTATTATTCCTTATCCTTAGGTGA.....GAACCTATA 592
173 sPheLysValIleMetSerIleGlyTyrSerAspSerGluAsnPheL 190
593 TGATGATATGAAAAAGAA.....ATTGATTATGTGATATAATA 633
190 yStIleIleIleLysAspLysLeuGlnAsnPheValAspSerSer 206
634 TTGAAGCTTTGATTTGATTTAGATGCTGATATTTAGCTGGGA 683
207 ValGluThrMetPheArgLeuGlyPheAspGlyIleAspLeuAspTyrGln 223
684 ACCACATGGGAGTTTACAACTTAATGAATTAATTTTCAATTAAT 733
223 uPheProGly.....AsnAsnGluSerGluProArgGlyT 235
734 ATATTAAATTAATTAATTAATTTAGAAAAAACTATCCGAGAAAAAGTTA 783
235 yTyrLeuLysLeuValArgMetLeuArgLeu.....LysLeu 246
784 ATTTCATTTTGTGCTTCATCAAAATGCTGCATTAATCATCGTTTCAGAGT 833
247 AsnSerLeu..... 249
834 TGCATCTTTTCTGAAGATGAAGAATCTCATATTAACATAAATTTTGT 883
250 .....GluSerGlnIlePheG 255
884 CTGAACAATAGAAACAATTAAGATTACATAGGCGACGCGATGTTA 933
255 LysAspGThrGluAspHisPheGlnLeuSerIleAlaIleProAlaPhe 271
934 TCAGCAGAGACTTTATTAATTTTATACGAAAGGAGAAATGA 983
272 LysAspLysLeuPheTyrLeuProIleThrGlnIleAspGlnTyrValAs 288
984 TCTTGATTTTCAACATCAATTA.....GAACTTA 1018
288 pTyrTyrAsnMetMetThrTyrAspTyrTyrGlySerTyrSerGlnThr 305
1019 CAAATCCAGATATAATGATAGATATGACTTATCCCATTTATAT 1062
305 hArgLys.....TyrHisSerAsnLeuPheSerGlu 314
1063 .....TTTGCTTAAATATTAACATC..... 1083
315 ThrGluLeuAsnGlnLysAsnPheAlaMetHisTyrMetIleAspArgPheGln 331
1084 .....ACAATCATATAGGT.....T 1099
331 yValAsnSerArgLysLeuValIleuGlyMetAlaIleTyrGlyArgSerP 348
1100 TTTTCATAGAACATACAGAGGTGATTTAGTCC.....GAAATATAA 1143
348 hHisIleLysAspAsnLys.....PheGluProPheAsnGlnAsnThr 362
1144 GAATTAATTAATTTAGGAAAAACAATACATATAATAAATCAAAATTA 1193
363 ValIleuLeuAsnLysIlePheLysGlyValGlyLysProThrLysGluI 379
1194 TAATAGGCGCATGTGT...ATAGGATATAGCATTTATTTTGAAGAAC 1240
379 eAspLysAlaAspLysLysGlyLysGlyTyrP..... 389
1241 AATTACCACTGATCATGATGATAGATATTTTCTTACAAATATTTGG 1290
389 ..... 389

```

```

1291 AAACATTTAAATCCTGAAGTACAAATCCAAAGACCTTACTATTAAGTGA 1340
390 .....ProTyrLysAsnLeuProLys..... 396
1341 AAACCCCTAGACCTGTAGCACATAGATGAATATGTTCCAGAGACTGTTA 1390
397 .....IleGlyThrIleGluGlnTyrAspProLysTyrValS 409
1391 TTCCA.....ACCATAGGATATATTTACAAACAC 1419
409 eAlaTyrCysPheAspLysAsnSerIlePheIleSerTyrAspAsn 425
1420 AATGATGCTATATGAAAACTAGATCTTATTAATCAATCATCACCCTGTGT 1469
426 ThrLysSerValLysThrLysAlaGluTyrValThrHisAsnLeuGln 442
1470 AGACGATATGAAATGGGACTTGCTGCAAGATATGCTATGAAAAA..... 1512
442 yGlyGlyPheThrPyrGlnSerCysGlyGluAlaTyrAlaAsnGlnSerA 459
1512 ..... 1512
459 rgSerLeuIleAsnAlaPheAsnGlnGlyLeuHisPheAsnValSerSer 475
1513 .....ATATGCGATGGGAAAGACGCCATTAATTAATTAACACTGA 1550
476 LysProSerIlePheGlnAspValArgValLysLysTyrTyrLeuAsnLys 492
1551 CTATTAAGAAAGCTCTATTAATTAATGAAAGGAGACCATATTTA 1596
492 sTyrGlyLysPheLysGlyPheLeu.....SerProTyrLeu 503
seq_name: sp_bacteria:O54328
seq_documentation_block:
ID 054328 PRELIMINARY: PRT: 563 AA.
AC 054328:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHITINASE (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-
DE ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).
CN CHIA.
OS Enterobacter sp.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=42895;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G-1;
RA Park J.R., Okamoto T., Yamasaki Y., Tanaka K., Nakagawa T.,
RA Kawamukai M., Matsuda H.;
RL J. Ferment. Bioeng. 84:493-501(1997).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-
CC ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC EMBL: U35121; AAB97779.1; -.
DR HSSP: P07254; 1CTN.
DR InterPro: IPR001579; Chitinase_2.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR002173; PFKB.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR SMART: SM00089; PKD; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
DR Glycosylase; Hydrolase.
SQ SEQUENCE 563 AA: 61017 MW: E455EC15B050AE37 CRC64;

```

alignment\_scores: quality: 148.50 length: 425

Ratio: 0.746 Gaps: 20  
Percent Similarity: 46.824 Percent Identity: 20.706  
alignement\_block:  
US-09-579-383-2 x 054328 ..  
Align seg 1/1 to: 054328 from: 1 to: 563

```

25 TTAATTTAGTATCCATCTGTATTCGCAAAATTCGAGAACCTTGAAAG 74
   |||:||||| ||| ||| |||: ||| ||| ||| ||| ||| |||
11 LeuLeuIleGlySerThrLeuGlySerSerAlaIleGlnAlaIleProG 27
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 AAAAAAATATTA.....AATTAATTCATGGGAATATACGGGAAA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 YLSPProThrIleAlaTrpGlyAsnThrLysPheAlaIleValGluVal 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 ATAAATAATTAACATCAACGGAATACATGAGCTTTTTCACATCTT 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 spGlnAlaIleThr.....AlaTyraAsnAsnLeuValLysVal 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 AAATCGAATATAGTAAATTTTGTAGAAATAGCATCTATTCGGAGAT 213
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 LysAsnAlaIleAspValSerValSerTrpAsnLeuTrpAsnGlyAsp 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 .GGGTGTAACCTGATGATTTACAAAAAATATATAA..... 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 aGlyThrThrAlaLysIleLeuLeuAsnGlyLysGluAlaTrpSerGly 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 .....AATATAATTAATAATGAT 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 roSerThrGlySerSerGlyThrAlaAsnPheLysValAsnLysGlyGly 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 AGA..... 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 ArgTyrlGlnMetGlnValAlaLeuGlyAsnAlaAspGlyCysThrAla 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 ..... 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123 rAspAlaThrGluIleValAlaAspThrAspGlySerHisLeuAlaP 140
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 .....AAATCAACCAAGCAAAATTTTACAGAGATTAATAAAGCAAC 312
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 roleuLysGluPro.....leuLeuGluLysAsnLysProTyLysGln 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 GGT.....ATTATAGCAGATACTATGTTTCATGAGCACTCAAG 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 AsnSerGlyLysValAlaGlySerTrpPheValGluTrpGlyValTyrl 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 TGATAGAGCAAAACATATGATGATTCAAACCCATGTCATTTTAT 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 Y.....ArgAsnPheThrValAspLysIleProAlaGlnAsnLeuThr 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
404 ATATTGCATTGCTCGCAT.....AATATGTTTATGAT 438
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 isLeuLeuTyrlGlyPheIleProIleGlyAsnGlyIleAsnAsp 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
439 .....GTATCTAGACCATTTAA 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 SerLeuLysGluIleGlnGlySerPheGlnAlaLeuGlnArgSerGyl 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 TGGAGAGCAAAAGATCTCTATTAGAAAACAC.....GGCTTAG 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 ngLyArgGluAspPheLysValSerIleHisAspProPheAlaLeuG 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
494 AATATGAACCTAGTATGATG..... 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 LuLysAlaGlnLysGlyValThrAlaTrpAspAspProTyrlLysGly 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 CTTATATCAATTAGAGTATCAGAAAAGTACGCCAGATTAATTTCT 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 PheGlyLysLeuMetAlaLeuLysGlnAlaArgProAspLeuLysIle 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 TTTATCTTGGTGGAGAAAC.....TATATGATAG 598

```

```

269 | |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 uProSerIleGlyLysThrPheLysSerAspProPhePheMetGlyA 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
599 ATATAGAAAAAGAAATGATATGTCGATTAATAATATGAACTGTTAAT 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 sPlyValLysArgAspArgPheValGlySerValLysGluPheLeuGln 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
649 GATTTGAT...TTAGATGCTGATATATGACTGGAA...CCACATGG 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGly 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
693 GAAGTTTAC.....AAGTTAATGAATTAATTTTCAATATATATA 736
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 YLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyLysThrTy 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
737 TTAATTAATTAATCTGTTAAGAAAACATATTCGGAAGAAAGTAAAT 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 alleuLeuMetLysGluLeuArgAlaMetLeuAspGln..... 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 TCAATTTCTGGTTCATCAAAATGCTGATTCATTCATGCGTTCCAGAG 836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 .....LeuSerAlaGluThrGly..... 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
837 ATCTTCTGTAAGATGACAGATCTCCATATACACTAAATTT...TTGT 883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 .....ArgLysTyrlGluLeu 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 CTGAACAATAGAAACAATAAAGATTAACATAGCGACGACGATGTTA 933
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 hrSerAlaIleSerAlaGlyLysAspLysIleAspLysValAla..... 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
934 TCACGACGAACCTTTATTAATTTTAAATACAGCAAAAGCAAAATAGA 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 .....TyraAsnValAlaGlnAsnSerMetAs 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 TCTGTATTTATTCAAACATACAT 1008
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 PHisIlePheLeuMetSerTyraSp 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: sp_invertebrate:Q25802
seq_documentation_block:
ID Q25802 PRELIMINARY; PRT; 960 AA.
AC Q25802;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FRAMESHIFT.
GN RPOD.
OS Plasmodium falciparum.
OX Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C10;
RX MEDLINE=96346169; PubMed=8757284;
RA Wilson R.J.M., Denny P.W., Preiser P.R., Rangachari K., Roberts K.,
RA Roy A., Whyte A., Strath M., Moore D.J., Moore P.W., Williamson D.H.;
RT *Complete gene map of the plastid-like DNA of the malaria parasite
RT Plasmodium falciparum.
RL J. Mol. Biol. 261:155-172(1996).
DR EMBL: X95275; CAA64574.1; -
DR InterPro: IPR00722; RNA_pol_A.
DR Pfam: PF00623; RNA_pol_A: 1.
SQ SEQUENCE 960 AA; 117987 MW; F41782D73607F35D CRC64;

```

alignment\_scores: Quality: 148.50 Length: 661  
Ratio: 0.488 Gaps: 43  
Percent Similarity: 45.991 Percent Identity: 22.995  
alignment\_block:

US-09-579-383-2/rev x Q25802 ..

Align seg 1/1 to: Q25802 from: 1 to: 960

```

1561 TTTCTTATAGTCAGTGTATATATAGGCGTTCATCCG... 1517
1561 PheTyrrhePheasnLysTyrrasnLeuLysIleuGluLysLeuLe 19
1516 .AATATTTTCATAGCATCTTTCAGCAAGTCCATTCATCTCTAC 1468
19 uillelePhe.....LysTyrrasnIleSerPheL 29
1467 ACCAGGTCAGATGATGATAAGATAGTTTCCATATACCATCTGT 1418
29 ysiIleuHisGluLeuLeuTyrrLeuGluTyrrIleuSerPhe... 43
1417 GTTGTATATATCCCTATGTTGGAATTAACGAGTCCGAAACATATCA 1368
44 .....LeuTyrrasnTyrrSerLeuasnIleLysAspSerasnPheI 58
1367 TCATATGCTGCTACAGCTTCAGGTTTTCAGTATAGTCTTTGG 1318
58 leTyrrLeuLeuIleLeuTyrrLys...AsnLysIleasnIleTyrrasn 73
1317 AGTTTGACTTCAGGATTTAAATGTTCCAAATATTTGTAAGAAAATAT 1268
74 AsnLysTyrrTyrrGluLeu.....LysAsnAs 82
1267 CTACATGATGATCATCCAGTTGGTATATGTTCTTC...ATAAT 1227
82 nTyrrIle.AsnValPheLeuasnTyrrTyrrTyrrLeuLysValIleasn 98
1226 AAATGCCATATCCCTATACCATCTGCCCTATATATTTGATTTTATC 1177
99 Lys.....IleGlnGluIleLeu..... 104
1176 ATGTATGTTTTCCTACCAATTTCTAATATCTTTATTTTCGGACTAA 1127
105 .....AsnAsnAsnLeuTyrrasnLysIleA 113
1126 ATCCACCTCTGTTATGTTCTAATGAAAAAACCCTATATGATGATGTTA 1077
113 snPro.....IleTyrrSerasn.....LeuPheLeu 121
1076 TATTTTAAACCAAAATATAA..... 1056
122 PhePheasnLysIleLysIleTyrrSerGlnLeuGlnIleuI 138
1055 ....TGGGATAAGTACATATCTACCATATATCTGGATTTGACTTTCTA 1010
138 eglTyrrLysGluTyrrIleSerasn...IleLysGluMetIleTyrrGlu 154
1009 AATTGATGTT.....TGAATAAATACAGATCTATTTCTCTCTGCT 966
154 ysrProValIleasnAsnTyrrIleasnGluLeuasnIleTyrrGluTyrrIle 170
966 ..... 966
171 LeuSerCysTyrrGlySerLysLysGluIleIleAspPheIleuLysIh 187
965 .....GTATTAATAATATTAATAAAGTCCGCTGATA 932
187 rAlaAspSerGluTyrrLeuThrLysGluLeuIleasnIleThrSerasn 204
931 ACATGCTGCTGCCCTA...TGTAAATCTTATTTGTTTCTATTTGTTCA 885
204 heIleIleLysGluLeuasnCysLysSerProPheIle...LeuLysTyrr 219
884 GACAAAATTTAGTGTATATGAGAT..... 858
220 IleLeuasnMetaspIleTyrrGluAsnIleIleLeuProLeuasnIleLe 236
857 ...TCTTCATCTTTACAGAAAGATGCACCTCTGAAACGATGATATAG 812

```

```

236 uArgPheLysIleLeuGlnAsnAsnIle..... 245
811 CAGCATTTGATGACCAAGAAATGAATTAACCTTTCTCCGGAATA... 765
246 .....LeuAsnLeuasnAsnGluTyrrPhe 253
764 .....GTTTTCCTAACAGTTAAT 745
254 IleTyrrThrLysasnThrTyrrIleThrLysTyrrIleLeuasnLysIleuLe 270
744 TAATTTATATATAT.....AATTGAAA..... 722
270 uAsnLeu.TyrrAsnArgArgAsnIleTyrrLeuasnIleLysSerValTyrr 286
721 .....AATTAATTCATTTAAGTTGTAACCT..... 695
287 LeuCysasnIleTyrrAsnAsnIleCysAsnThrCysLeuasnTyrrLysGlu 303
694 .....TCCCATGTGCT.....TCCCATG 677
303 nLeuTyrrLysTyrrasnLeuGlnHisIleGluValIleSerSerGlu 320
676 CAATATCTACACCATCTAAATCAAAATCATTAACAGCTTCATATTTTA 627
320 laIleSerGluProSerThrGlnMetValLeuArgThrPheHis..... 334
626 TCACATATATCAATTTCTTTTATATCTATCATATATAGTTTCTCCACC 577
335 .....AlaSerSerIleLeu..... 339
576 TAAGATATAAAGAAATTAATACATCTGACGCTACTTTTCTGATACGTCTAA 527
340 .LysAspLysPheasnPheasnLysTyrrLeuIleTyrrLysIleTyrrLeu 356
526 TTTCATTAAGCATCATACCA.....TAGGTTTATTTCTTAAG 489
356 TyrrLysLeuasnIleAsnLysIlePheLysLeuIleIleAsnPheLysLys 372
488 CCGGTTTCTTATATAGGAATCTTTGCTTCCTCAATTAAGTGTACATATAC 439
373 TyrrIleasnIleLysPheasnLeuIlePheLeuMetasnLysIleLeuTy 389
438 ATCATAT...ACATATTTAATGCGAG.....CAATGCAA 407
389 rAsnTyrrAsnAsnIleLeuPhe.GluTyrrLysTyrrIleuGlnAsnGln 405
406 TATATAAATTTGACACCATGCGTTGGAATCATATATGTTTGGCTCTA 357
406 TyrrIleLysCysasnPheIleTyrrAsnSerIleSerLysAsnPheLysTy 422
356 TCACCTGACTGTTCCATGAACCATATGATCTGATATACCT... 311
422 rAsnLeu.....AsnAsnIleIleIleLysTyrrLeuAsn 434
310 .....GTTTCTCTTTT...TATACCTCTCTA..... 287
434 snValIleLysTyrrTyrrAsnTyrrSerasnIleGlnLeuLeuIleLysasn 450
286 .....AATTTGCTTGCGATTTTCTATCATTTTATTTATTTATTT 246
451 IleHisasnLysTyrrIleLeuTyrrAsnIleTyrrThrTyrrLeuTyrr 467
245 TTTATTTTCTTTTGTAAATCTAGATTAACAC..... 215
467 TyrrHisIleLysPheTyrrasnLeuTyrrasnLysGluIleIleLeuAsn 484
214 .....CATCTCCGCAATAGATCAT 194
484 snAsnAsnAsnLysTyrrAsnValIleTyrrPheLeuIleasnTyrrPheasn 500
193 ATTCTACAAATTAATTAATTTGATTTAAGAT.....GT 159

```

```

501 LeupheserAsnTYR TYR Lys Ile TYR Asn Asn sntYr Asn Phe Il 517
158 GAAAGACATCATGATATTCCTGTTGATGATTTATTTTA.....TT 115
517 eAsnSerAsnTYR TYR Phe Lys Lys Met Asn Phe Ile Leu Lys Asn Ph 533
114 TTCCCGATTT...ATTCGAATGAAATTTATATTA.....TTTTTC 74
533 eAsnAsn Ile Glu Ile Leu Asn Lys Leu Phe TYR Val Asn Ile Phe I 550
73 CTTCAAGCTCTCGAATTTGCAGATACAGATGATACATATATAA 24
550 l eTYR TYR Lys TYR Glu Lys Lys Leu Phe Ile TYR Leu Asn Ile Leu 566
23 AATATGATATTTTAAATTC 3
567 Asn Ile Ile Ile Lys Lys TYR 573

seq.name: sp_archaea:Q9UWR7

seq_documentation_block:
ID Q9UWR7 PRELIMINARY; PRT; 1215 AA.
AC Q9UWR7;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE CHITINASE.
GN PK-CHTA.
OS Pyrococcus kodakarensis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-20049967; PubMed-10583986;
RA Tanaka T., Fujiwara S., Nishikori S., Fukui T., Takagi M., Imanaka T.;
RA "A unique chitinase with dual active sites and triple substrate
binding sites from the hyperthermophilic archaeon Pyrococcus
kodakarensis KOD1."
RT Appl. Environ. Microbiol. 65:5338-5344(1999).
DR EMBL; AB024740; BAA88380.1; -.
DR HSSP; P07103; IAIN.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR001579; Chitinase_2.
DR InterPro; IPR003610; Chitin_bind3.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00553; CBD_2; 2.
DR Pfam; PF00704; Glyco_hydro_18; 2.
DR SMART; SM00495; ChitD3; 1.
DR PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
SQ SEQUENCE 1215 AA; 134266 MW; 7A3997EBE7A86C0 CRC64;

alignment_scores:
Quality: 148.00 Length: 647
Ratio: 0.576 Gaps: 28
Percent Similarity: 39.722 Percent Identity: 17.311

alignment_block:
US-09-579-383-2 x Q9UWR7 ..

Align seg 1/1 to: Q9UWR7 from: 1 to: 1215

160 CATCTTAATCGAATAAGTAAATTTGTAGAAATGATGATCTATTTCGG 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 HisLeuLysAlaAsnThrThrTYRTYRGLYValValProValLeuAl 155
210 AGATGGGTGCTACTCTGAAATTCAAAAAATAATAATAATAATAATAA 259
:|||||:
155 aAspGly..... 157
260 ATGATAGAAATCACAAGCAAAAT.....TTAGAG 291
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 ..SerAlrGlySerProSerAsnValLeuAlaIleThrThrProLeuGlu 173

```

```

292 GAGTATATAAAAAAGAAACAAGTATTATAGCAGATACATATGCTCATG 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 ProTYrArg.....ValIleValTYR ILeSerTr 184
342 GAACAGTCAGAGCATAGACCAAAACATATGATTCACAAACCATGCG 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 pGlyArgTYrAla.....ArgLysPheTYR ValSerAspLeuProTrpG 199
392 TGTCATTTTATATATGCAATTCGCAATTAATGATATATATATGATGTA 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 LuLysValThrHisValAsnTYRAlaPheLeuAspLeuLysGluAspGly 215
442 TGTACACATTTAATGAGACAAAGATTCCTATTAGAAAACACGCGCTT 491
:::|||||:
216 ThrValAlaPhe..... 219
492 AGAATATGAAACCTATGCTATGATGCTTAAT.....GAAATGACGTA 535
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 ...TYRAspThrTYRAlaAspProLeuAsnLeuGluAlaMetLysGluT 235
536 TCAGAAAGTACGTCAGATGATATATTTCTTTATCTTAGTGAGAA 585
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 YrLysAlrGlySTYRProAlaValLysValLeuIleSerValGlyTYR 251
586 ACC.....TATATGATGATATAGAAAAAGAAAT 614
|||:|||||:
252 ThrLeuSerLysTYR PheSerValValAlaAlaAspProAlaLysArgL 268
615 TGATTATGATGATAAATATTGAAGCTGTTAAATGATTTGATGATAGAT 664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 naArgPheAlaGluThrAlaIleGluIleLeuArgLysTYRAsnLeuAspG 285
665 GTGTAGATATTTGACTGGAA.....CCACATGCGAAGTTTCAAC 705
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LysLeaPheIleAspTrpLeuTYRProLysGlyGlyMetAlaGlyAsn 301
706 TTTAAATTAATTAATTTTCAATTAATTAATTAATTAATTAATTAATGTT 755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 TYrGluSerProAspAspGlyLysAsnPheValLeuLeuLysAspLe 318
756 AAGAAACATATTCGGAAGAAAGTAAATTTCAATTTGCTGCTCACAA 805
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
318 uArgGluAlaLeuAspLysAla..... 325
806 ATGCTGATATATCATCGCTTCAAGAGCTGCATCTTCTGAAGATGAA 855
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 .....AlaLysGluAsp 329
856 GAATTCATATTAACACTAAATTTTGTCTGACAAATAGAAACAATAA 905
|||||:|||||:
330 HisLysAspTYR..... 333
906 AGAATTACATAGGCGACAGCAGATGTTACGACGAGACTTTTATATA 955
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334 .....LeuLeuThrAlaAlaThrProAlaAsp 343
956 TTTTATATA.....GCAAGAGCAAAATAGAT 984
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 roValLysAlaGlyArgIleAspTrpValGluAlaSerLysTYRLeuAsp 359
985 CTGTGATTTATCAACATAC..... 1005
:::|||||:|||||:
360 SerIleAsnIleMetThrTYRAspTYRHisGlyAlaIleTrpGluThrIleTh 376
1006 .....AATTTGAACAATACAAATC 1024
|||||:|||||:
376 rGlyHisLeuAlaProLeuTYR CysAspProAsnAlaProTYRThrAspG 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1025 CAGATATA.....ATGTAGATATGACTTA 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 LuAsnValLysTYRHisPheCysValAsnTYRThrValGluIntPrpTYrIle 409

```





```

107 ArgTyrGlnMetGlnValAlaLeuCysAsnAlaSpGlyCysThrAlaSe 123
267 ..... 267
123 RasPalaThGluIleValAlaAlaSpThrAspGlySerHisLeuAlaP 140
268 ....AATACCAACAAGCAATTTTAGAGAGTATAAAAAGAAACAA 312
140 roleuYsGluPro.....LeuGluIuYsAsnYsProTyrLysGln 154
313 GGT.....ATTATAGCAGACTATGTCATGTCAGAACAGTCAAG 353
155 AsnSerGlyYsValValGlySerTyrPheValGluTyrGlyValTyrG 171
354 TGATAGACAAACATATGATTCATTCACCAACCATGTCGCAATTTAT 403
171 Y.....ArgAsnPhetHrValAspLysIleProAlaGlnAsnLeuTrr 186
404 ATATTCGATTTGCTGCAT.....AATATGTTATATGAT 438
186 IStleuTyrGlyPheIleProIleCysGlyGlyAsnGlyIleAsnAsp 202
439 .....GATCTACACCATTTAA 455
203 SerLeuYsGluIleGlyGlySerPheGlnAlaLeuGlnArgSerCysG 219
456 TGGAGACAAAGATTCCTATTAGAAAAC.....GGCTTAG 493
219 nglyArgGluAspPheLysValSerIleHisAspProPheAlaAlaLeuG 236
494 AATATGAACCTATGTATGATG..... 516
236 InYsAlaGlnIuYsGlyValIThrAlaTrrAspAspProTyrLysGlyAsn 252
517 CTTAATCAATTTAGACGTATCGAAGAGTACGACGATGTAATTTATCT 566
253 PheGlyGlnMetAlaLeuIuYsGlnAlaHisProAspLeuLysIleLe 269
567 TTTATCCTTAGGTGAGAAAC.....TATATGATAG 598
269 uProSerIleGlyLyrTrpHrLeuSerAspProPhePheMetGlyA 286
599 ATATAGAAAAGAAATGATTTATGCTAATAAATTTGACCTGTTAT 648
286 spLYsValIuYsArgAspArgPheValGlySerValIuYsGluPheGln 302
649 GATTTGAT...TTAGATGCTGATATGATGACGGGAA...CCACATGG 692
303 ThrTrpLysPhePheAspGlyValAspIleAspTrpGluPheProGlyG 319
693 GAAGTTTAC.....AACTTAATGAATTAATTTTCAATTTATATA 736
319 YLysGlyAlaAsnProAsnLeuGlySerProGlnAspGlyGluHrYtrV 336
737 TTTAATTAATTAACGTGTAAAGAAAACATTCGGAAGAAAAGTTAAT 786
336 AlLeuLeuMetLysGluLeuAlaGlnAlaMetLeuAsnGln..... 348
787 TCAATTTCTGTCATCAATGCTGCATTAATCATGCGTTTCAGAGATTGC 836
348 ..... 348
837 ATCTTTCTGTAAGATGAGAAATCTCATATTAACATAATTTTGTCTG 886
349 .....LeuSer. 350
887 AACCAATGAACAATAAAGATTACATAGGACGACGACGATGTTATCA 936
351 .....ValGluThrGlyArgGlySerGluLeuHrSerAlaIleSerAla 365
937 GCAGGAACCTTTTATTAAT.....ATTTTAAATACAGCAAGGAGAAAT 980
:::

```

```

366 GlyYsAspLysIleAspLysValAlaTyrAsnValAlaGlnAsnSerMe 382
981 AGATCTGTATTTATTTCAACATCAAT 1008
382 tAspHisIlePheLeuMetSerTyrAsp 391

```

seq\_name: sp\_invertebrate:096206

```

seq_documentation_block:
ID 096206 PRELIMINARY; PRT; 1817 AA.
AC 096206;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE PREDICTED SECRETED PROTEIN.
GN PPH0565W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteh H., Carnocci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petter M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser G.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AB001403; AAC71902.1;
SQ SEQUENCE 1817 AA; 217996 MW; D1812785960E0BCE CRC64;

```

alignment\_scores:

```

Quality: 144.00 Length: 558
Ratio: 0.514 Gaps: 31
Percent Similarity: 50.179 Percent Identity: 20.968

```

alignment\_block:

US-09-579-383-2 x 096206 ..

Align seg 1/1 to: 096206 from: 1 to: 1817

```

64 ACCTTGAAGGAAATAATTAATTAATTCATTCGGAATAATACGGCA 113
111 ::::::::::::::::::::
355 ThrLysGlnGlyLysHisAspMetAsnAsp.....Tyr 365
114 AATATTAATAAATCAATCATCAACGGAATACATGATCTTTTCACATC 163
111 ::::::::::::::::::::
365 rAsnMetAsnLysLysAsnAsnMetAspIleAsnIleThrIleAsnAsnA 382
164 TTAATATCAATTAATAGTAATTTTGTAGAATATGATCTTTATTCGGGAGAT 213
382 snAsnAsnAsnAsnAsnAsnLysIle.....TyrAsnAspAsn 394
214 GGGGTAACTCTAGCAATTAACAAAATAATTAATTAATTAATAA..... 258
111 ::::::::::::::::::::
395 ThrLeuAsn.....ValTyrAsnAsnSerTyrAsnIleHisSerAsnH 409
259 .....AATGATAGAAATAATCACCAGCAAAATTTTAGAGGAGTATATAA 301
111 ::::::::::::::::::::
409 sLeuMetAsnAspLysArgLysAsnAlaGlnValLeuGluLysHisLeu 426
302 AAAGGAACAAGATATATAGCAGAGATACATAGTTTCATGAGACAGTCAA 351
111 ::::::::::::::::::::
426 ys.....MetLeuCysAspAsnPh.....PheAsnLeuGlu 436
352 GGTATAGACCAAAACATATGATTTGATTAACCAACCAATGTCGCAATTTT 401
437 GluPheTyrSerSerAsnIleIleIleAsnAsnMetAspIleIuYtrH 453
402 ATATATTCATTTGCTGCATTAATATGTTATATGATGTTATGACCAT 451

```



```

453 rtyrasyrprphe.....ileutyrglu.....lyscysp 464
452 TTAATGGAAGACAAAGATTCTATTAAAGAAACACGCTTAGAATATGAA 501
464 heluProIleGluArgIleVal.....HisValAsnTyrMet 476
502 ACCATTGATGATGCTTATATGAAATTAACGATATCAGAAAGTACGTC 551
477 LysTyr...LeuTyrLysAsnAsnGluArgLysAsnLysLysIleArgLy 492
552 ACATGTAATATTCTTTATCCTTAGCTGAGAACCTATATGATGATGATA 601
492 sPheLeuIleThrLeuLeuGlu.....TyrSerArgAspI 504
602 TAGAAAAAAGAAATGATATTGATGATTAATATG...AAGCTTGTTAT 648
504 leAsnPhelHisPhePheIlePheAsnLeuIleLeuTyrLysCysLysAsn 520
649 GATTTTGATTTAGATGCTGATACATATTGACTGGAAACACATGGAGATT 698
521 GlupheProCysSerIlePheGluLeuHisIleSerGlnTyrLeuTyrPh 537
699 TTAGC...AAGTAAATGATTAATTTTCAATATATTAT... 738
537 ePheValLysLeuAsnGluLeuAsnLysLysAspIleTyrIleTyrTyr 554
739 .....AATTAATTAACTTGTTAGAAAAAAGCT 765
554 heAsnAsnPhelTyrGlnAspMetIleIleTyrPheSerArgLysAla 570
766 ATT...CCGGAAGAAAGTTAATTTCAATTTCTGTTCAATCAATGCTGC 812
571 PheTyrProTyrPgl.....ThrAsnValGluGlnGlnLysLysGlnTh 585
813 ATTATCAGCGCTTCAGAGTGCATCTTCTGTAAGATGAAGATCTC 862
585 rleuSerTyrIle..... 589
863 CATATTAACACTAAATTTTGTCTGAACAATAGAAACAATTAAGAATTA 912
590 ..TyrAsnAspLys.....IleLysLysAsnLysLys... 599
913 CATAGCGACGACGATGTTATCAGCAGAGACTTTATTAAT..... 954
600 AsnAsnSerTyrTyrGluMetAsnAsnThrTyrMetAsnGluHisG1 616
955 .....ATTTTAATACAGCAAGAGAAAGAAATAG 982
616 YTyThrAspIleGluAsnGluArgLeuAsnLysLysAsnLysArgLeuA 633
983 ATCTTGATTTATTCAACATACATTTAGAAACTACAAATCCAGATATA 1032
633 snVal.....ArgLysArgThrAsnThrLeuAspAspIle 644
1033 ATGCTAGATGATGTTATCCCATTTATATTGCTTAAATATTAACAT 1082
645 lleValSer.....AspHisG1LysAsnSerTyrAspLysTyrAsnTh 658
1083 CACATCATATTAGTTTTTCATTAGAACATPACAGAGGTGATTTAGTC 1132
658 rSer.....LysHisAsnArg..... 663
1133 CCGAAATTAAGAAATTTATTAGAAATTTGTAGAAAAACAATACATGATAA 1182
664 ....ArgLysAsnHisIleAsnGluMetLysLysLysGlnAsnAsnLys 678
1183 AATCAAAATATATATAGGCAAGATGCTATAGGATATGCAATTTATTAT 1232
679 LysLysAsnThrLeuPheValAspG1LysAspMetGluGlyIleGlyLy 695
1233 GAAAGAACAAATTACCACTGATCATTCGATGATGATATTTTCTTACAA 1282

```

```

695 sGluLysGluLysGluAsnLysAsnMetAsnAsnIlePheTyrAsnA 712
1283 AT.....ATTGGAACATTTAAATCCTGAA 1308
712 snSerTyrSerAsnIleAsnAsnSerSerTyrSerAsnIleAsnAsnAsp 728
1309 GTACAAACTCCAAAGACCTTACTATTAATGAAACCCCTGAGACTGTAG 1358
729 IleTyrSerValAspAsnMetThrSerValAsnAsnThrLysTyrValSe 745
1359 CACAATAGATGAATATGTTCCAGGACTCGTTATTCACCATGAGGATAT 1408
745 rGlyValProSerTyrAlaHisValIleuIleAsnLysGlnValAsnGluT 762
1409 ATTACAAA.....CACAAATGATGCTATATGAAAACTAGATCT 1446
762 YTyTyrGlnGlyLeuProAsnTyrAsnAsnMetMetIleLys...GlySer 777
1447 TATTCAATTCATGACACCTGCTGTAGACAGATATGAATGGGACTTGGTCA 1496
778 HisIleIleAsnGluLeuProLysAsnAsnTyr..... 788
1497 AGTATGCTATGAAAAAATATGCGATGGGAAAGCAGCCATTAAT..... 1539
789 ....IleTyrGluAsnAsnTyrIleGlyGln.....AsnTyrLeuMet 802
1540 .....TATAACACTGACTATTAAGAAAGCTCTATTATTTATATAG 1578
802 hrAsnProLeuTyrAsnLysGluThrLysAspIlePheTyrThrIleTyr 818
1579 AAGGGGAACATATTTAATTA 1602
819 Lys.....TyrLeuPheLys 823

```

